Mesh independent superlinear convergence of some iterative methods for elliptic problems

István Antal

PhD Thesis

Supervisor: János Karátson
Associate Professor, PhD

Mathematical Doctoral School

Director: Professor Miklós Laczkovich
Member of the Hungarian Academy of Sciences

Doctoral Program: Applied Mathematics

Director of Program: Professor György Michaletzky
Doctor of the Hungarian Academy of Sciences

Department of Applied Analysis and Computational Mathematics
Institute of Mathematics
Etvs Lornd University, Faculty of Sciences
2012

CONTENTS

ACKNOWLEDGEMENT

I would like to thank my supervisor for his help and guidance, my professors whose lectures inspired me to pursue mathematics further. And of course my family and friends for their support.

OVERVIEW

The theory and numerical treatment of partial differential equations (PDE) has long been in the center of interest because of the overwhelmingly wide spectrum of its applications in the natural sciences, most prominently in physics, chemistry and engineering. Soon it turned out that even the simplest differential equations doesn't have solutions in the classical sense so there was a need to redefine what one means as a solution of the equation. This lead consideration lead to the theory of distributional and weak solutions. As for elliptic equations the proper way of the definition of a weak solution lead to the theory of Sobolev spaces, thus solutions are seeked in a much broader set of functions. Also there are virtually no real-life equation that has as solution that has a solution presentable in a close form thus there was, and there is to our days, need for methods to aquire approximate solutions. The presentation of the finite element method was a major breakthrough in the field of numerical solutions to PDEs. One of its power is the theory of weak solutions of PDEs that it can rely on.

PDEs the equation approximated by a finite dimensional algebraic equations. Since linear differential operators are unbounded and they are approximated by finite dimensional matrices on can only expect ill-behaviour of the occuring matrices. That is the condition number of the constructed matrices tends to infinity as the mesh gets finer. Working with ill-conditioned matrices is usually very problematic because of the inaccuracy it adds during a solution provess. One way of weakening this problem is method of preconditioning. In our case a reasonable chosen preconditioner will ensure that the occuring matrices have uniformly bounded condition numbers independent of the finement of the mesh.

In this thesis we use a preconditioning technique that has the above property. The preconditioner will be the discretization of the main part of the elliptic equation, that is the discretization of the Laplacian, for finite element method it is the stiffness matrix. Although with preconditioning of this stiffness matrix we have to solve an equation that is itself ill-conditioned, but here this is the only ill-behaved element so we can choose an appropriate method to control the ill-behaviour.

We consider both linear and nonlinear classes of elliptic PDEs. The nonlinear equations are solved by a proper chosen variant of the Newton method, we obtain mesh independent superlinear convergence showing that under reasonable assumptions the

derivative of the occuring function bears Lipshitz or a weaker local-Lipschitz condition. The linear equation and the derived linear subproblems of the proposed Newton method for the nonlinear problems we consider will be seen as perturbations of the identity. For these kind of operators we use well-known convegence estimates for the conjugate gradient method. That is we show that by solving these equations with the conjugate gradient method the error tends to zero superlinearly.

The structure of this thesis is the following. In the chapter Preliminaries we summarize the necessary theoretical results that we will use of the fields of functional analysis, partial differential equations and iteration methods. In the rest of the thesis we expound our results following the timeline of our publications. In the end of each section we present numerical testresults to support our convergence results.

In the Chapter 1 we consider linear and nonlinear symmetric ellitic systems. The convergence estimate for the solution of the linear equation (in the nonlinear case the linear subproblem) is obtained via the notion of the Hilbert-Schmidt operator. The key of this result is that the inverse of the Laplacian is of Hilbert-Schmidt class.

In the Chapter 2 we consider only nonlinear elliptic equations. In the first section nonsymmetric systems and in the second section a class of interface equations are considered. In this section we replace notion of Hilbert-Schmidt norm and use a convergence estimate that analyzes the eigenvalues of the occuring operators. The key tool to achieve this is the Courant-Fischer min-max principle. This will not only give better estimates but the validity of existence of superlinear convergence could be broaden. In both cases we give explicit order of convergence of the solution method using Gelfand numbers which is a tool to measure the compactness of operators between Banach spaces.

In the Addendum we give a short comparison and also a small improvement of the convergence estimates that have been proven in the first two chapters.

1. PRELIMINARIES

In this section we summarize the facts that will be used, and also the corresponding notations.

1.1 Functional analysis

In this section, if not noted otherwise, X, Y will denote Banach spaces, H denotes a real Hilbert space. Furthermore $B(X,Y)$ and $B(X)$ denotes the space of bounded linear operators $X \to Y$ and $X \to X$ respectively. The spectrum of A will be denoted by $\sigma(A)$. The contents of this section can be found as whole or as part in [14, 41, 37].

1.1.1 Weak solutions

Definition 1.1. Let H be a Hilbert space and $A : D(A)(\subset H) \to H$ be a densely defined opertator.

- If B is also a densely defined operator on H then we say that B extends A, denoted as $A \subset B$, if $D(A) \subset D(B)$ and on the set $D(A)$ they coincide. This is trivially a preorder.
- The adjoint of A denoted as $A^* : D(A^*)(\subset H) \to H$ is the largest operator (w.r.t. the '$\subset$' relation) among the operators $\{B : D(B)(\subset H) \to H\}$ that has the property
$$\langle Ax, y\rangle = \langle x, By\rangle,\ x \in D(A), y \in D(B).$$
- A is symmetric if $\langle Ax, y\rangle = \langle x, Ay\rangle$, $x, y \in D(A)$, or shortly $A \subset A^*$
- A is selfadjoint if $A = A^*$.
- A symmetric operator A is positive/strictly positive/uniformly positive if
$$\begin{aligned} \langle Ax, x\rangle &\geq 0, \\ &> 0 \quad (x \neq 0), \\ &\geq m\|x\|^2 \text{ for some } m > 0 \end{aligned}$$
respectively for each $x \in D(A)$.

Definition 1.2. Let S be a uniformly positive operator on the Hilbert space H. Then $\langle x, y\rangle_S$, $x, y \in D(S)$ defines an inner product. The completion of this inner product space is the energy space H_S, the norm of this space is denoted by $\|\cdot\|_S$. Because of its construction it can be continuously embedded into H.

Remark 1.3. If S is bounded, then it can be extended to an operator on $B(H)$, and the norm $\|\cdot\|_S$ is equivalent that of H.

Definition 1.4. Let S be a uniformly positive selfadjoint operator on the Hilbert space H. Let $f \in H$, then the weak solution $u \in H_S$ of the equation $Su = f$ satisfies

$$\langle u, v\rangle_S = \langle f, v\rangle,\ v \in H_S$$

Theorem 1.5. *A positive operator extends to a selfadjoint operator (not neccessary uniquely). The selfadjoint extension constructed via the above mentioned energy space is the Friedrichs extension. A positive operator is selfadjoint if and only if it is surjective.*

Proposition 1.6. *Let S be a uniformly positive selfadjoint operator on the Hilbert space H. Then for each $f \in H$ the equation $Su = f$ has a weak solution. If S is only symmetric, then using the Friedrichs extension method mentioned in the previous theorem we are lead to a selfadjoint extension $\widehat{S}$. The domain of $\widehat{S}$ is defined as*

$$dom\,(\widehat{S}) = \left\{x \in H\ :\ \exists y \in H \text{ such that } \langle \widehat{S}x, z\rangle = \langle y, z\rangle \quad (\forall z \in H)\right\}$$

1.1.2 Compact operators

Definition 1.7. An operator $C \in B(X, Y)$ is compact if it maps bounded sets to totally bounded sets.

Proposition 1.8. *If in the above definition the Banach space X is reflexive, specially a Hilbert space, then an operator is compact if and only if it maps weakly convergent series to convergent ones.*

Theorem 1.9 (Riesz). *Let X be a complex Banach space and a compact operator $C \in B(X)$.*

- *The spectrum of C consists purely of eigenvalues, and possibly the value 0.*
- *The dimension of an eigenspace corresponding to an eigenvalue is finite.*
- *The eigenvalues has at most one accumulation point, that can only be the value 0.*

- *If the operator is selfadjoint then the same statements hold, since then all of its eigenvalues are real.*

Proposition 1.10 (Courant-Fischer min-max principle)**.** *Let $C \in B(H)$ a compact selfadjoint operator with eigenvalues $\lambda_1 \geq \lambda_2 \dots$ with multiplicity. Then we have the formula*

$$\lambda_n = \min_{H_{n-1} \subset H} \max_{x \perp H_{n-1}} \frac{\langle Cx, x \rangle}{\|x\|^2},$$

where $\dim(H_{n-1}) = n - 1$.

Definition 1.11. The Hilbert-Schmidt norm of an $A \in B(H)$ equals

$$|||A|||^2 = \sum_i \|Ae_i\|^2, \tag{1.1}$$

where $(e_i) \subset H$ is an arbitrary orthonormal basis.

Proposition 1.12. *The operators that have finite Hilbert-Schmidt norm are called Hilbert-Schmidt operators and they form a Hilbert space denoted by $B_2(H)$, where the inner product is induced by the norm $|||\cdot|||$. All the operators in $B_2(H)$ are also compact, and furthermore we have*

$$|||A|||^2 = \sum_i \lambda_i(A^*A).$$

where $(\lambda_i(A^*A))$ are the eigenvalues of the operator A^*A with multiplicity, they are also called the singular values of A.

Remark 1.13. The Hilbert-Schmidt norm is a natural extension of the Frobenius norm of matrices. Indeed for an operator on a finite dimensional space the right hand side of the formula (1.1) is the trace of A^*A which equals the Frobenius norm.

The following is a fairly straightforward corollary of the min-max theorem.

Remark 1.14. Let $A, B \in B_2(H)$ positive operators. If $A \leq B$ then for their ith largest eigenvalue we have $\lambda_i(A) \leq \lambda_i(B)$, this implies of course $|||A||| \leq |||B|||$.

1.2 Partial differential equations

In this work $\Omega \subset \mathbb{R}^d$ will always denote a bounded domain with sufficiently smooth boundary. Most of the contents of this section can be found in [1].

Definition 1.15 (L^p spaces)**.**

· Let $1 \leq p < \infty$. The collection of functions $f : \Omega \to \mathbb{R}$ such that

$$\|f\|_{L^p(\Omega)} = \|f\|_p = \left(\int_\Omega |f|^p \right)^{1/p} < \infty$$

endowed with this norm form a Banach space denoted by $L^p(\Omega)$. In the case of $p = 2$ we have a Hilbert space.

· Let $p = \infty$. The collection of $f : \Omega \to \mathbb{R}$ functions such that

$$\|f\|_{L^\infty(\Omega)} = \|f\|_\infty = \operatorname*{ess\,sup}_\Omega |f| = \inf\{M \geq 0 : |f| \leq M \text{ a.e.}\} < \infty$$

form a Banach space denoted by $L^\infty(\Omega)$.

Theorem 1.16 (Hölder inequality). *Let $1 \leq p, q \leq \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$ (p and q are called conjugate exponents then). If $f \in L^p(\Omega), g \in L^q(\Omega)$ then we have*

$$\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q.$$

Among the rich family of generalizations of the Hölder inequality we will use the following one.

Proposition 1.17. *Let $1 \leq p_1, \ldots, p_k \leq \infty$ such that $\sum \frac{1}{p_i} = 1$. If for $i = 1, \ldots, k$ $f_i \in L^{p_i}(\Omega)$ then we have*

$$\left\| \prod_i f_i \right\|_1 \leq \prod_i \|f_i\|_{p_i}.$$

Definition 1.18 (Sobolev spaces).

· **Integer Sobolev spaces**

Let $1 \leq p \leq \infty$ and $k \in \mathbb{N}$. The collection of $f : \Omega \to \mathbb{R}$ functions such that the distributional derivatives $D^\alpha f$ are in $L^p(\Omega)$ for $|\alpha| \leq k$ and

$$\|f\|_{W^{k,p}(\Omega)} = \|f\|_{k,p} = \left(\sum_{|\alpha| \leq k} \|D^\alpha f\|_p^p \right)^{1/p} < \infty$$

endowed with this norm form a Banach space denoted by $W^{k,p}(\Omega)$. If $p = 2$ we have a Hilbert space denoted by $H^k(\Omega)$, and if $k = 0$ we have $W^{0,p}(\Omega) = L^p(\Omega)$.

· **Fractional Soblev spaces**

Sobolev spaces $W^{p,s}(\Omega)$ can be defined for arbitrary $s \geq 0$ equivalently via Fourier multipliers, the theory of interpolation of Banach spaces or using the Slobodeckij seminorm. For details we only refer to the monographs cited above.

· **The space $H_0^1(\Omega)$**

Let $C_0^\infty(\Omega)$ denote the collection of functions whose support is compact and contained in Ω. The closure of this subspace $C_0^\infty(\Omega) \subset H^1(\Omega)$ is denoted by $H_0^1(\Omega)$.

- **Soblev spaces on the boundary**

  Sobolev spaces $W^{p,s}(\partial\Omega)$ can also be defined for the boundary $\partial\Omega$ for sufficiently smooth domains using the partition of unity on an open cover of $\partial\Omega$. For details we refer again to the monographs cited above.

- **Trace theorem**

  Let $1 \leq p < \infty$ and Ω be sufficiently smooth. Then the trace map

  $$\mathrm{tr} : C^\infty(\overline{\Omega}) \longrightarrow C^\infty(\partial\Omega) \ : \ f \longmapsto f\big|_\Omega$$

  extends to a surjective bounded operator

  $$W^{1,p}(\Omega) \longrightarrow W^{1/q,p}(\partial\Omega), \text{ where } q \text{ is the conjugate pair of } p.$$

Theorem 1.19 (Properties of Sobolev spaces)**.**

- **Friedrichs-Poincaré inequality**

  There exist $m > 0$ (depending on Ω) such that

  $$\|\nabla f\|_2 \geq m \cdot \|f\|_2 \text{ for all } f \in H_0^1(\Omega)$$

- **Equivalent norm on $H_0^1(\Omega)$**

  The norm $\|f\|_{H_0^1(\Omega)} = \|\nabla f\|_2$ is equivalent with the inherited norm $\|f\|_{H^1}(\Omega)$.

- **Embeddings**

  - *Sobolev embedding theorem*

    Let the boundary of Ω be sufficiently smooth and let $r > s \geq 0$ and $1 \leq p < q \leq \infty$ satisfying $(r-l)p < n$ and $\frac{1}{q} = \frac{1}{p} - \frac{r-s}{n}$. Then the following embedding is bounded

    $$W^{r,p}(\Omega) \hookrightarrow W^{s,q}(\Omega),$$

    i.e. for $f \in W^{r,p}(\Omega)$

    $$\|f\|_{s,q} \leq \|f\|_{r,p}.$$

  - *Rellich-Kondrachov embedding theorem Let the boundary of Ω be sufficiently smooth and let $1 \leq p \leq \infty$ and $s \geq 0$.*

If $r > s \geq 0$ *and* $r - \frac{n}{p} > s - \frac{n}{q}$, *then the following embedding is compact*

$$W^{r,p}(\Omega) \hookrightarrow W^{s,q}(\Omega).$$

$H = L^2(\Omega)$ and $S : C_0^\infty(\Omega)(\subset H) \to \mathbb{R}$ be defined as $Sf = -\Delta f$, this is a densely defined uniformly positive operator because of the Friedrichs-Poincaré inequality, hence it has the Friedrichs extension $\widehat{S}$, the associated energy space $H_S = H_{\widehat{S}}$ coincides with the Sobolev space $H_0^1(\Omega)$. Hence the weak solution of the homogenous Dirichlet problem

$$\begin{cases} -\Delta u &= f \in L^2(\Omega) \\ u|_\Omega &= 0 \end{cases}$$

is defined as $u \in H_0^1(\Omega)$ satisfying

$$\int_\Omega \nabla u \cdot \nabla v = \int_\Omega fv \quad \forall v \in H_0^1(\Omega).$$

Theorem 1.20 (Kadlec[23]). *If* $n = 2, 3$ *and* $\Omega \subset \mathbb{R}^n$ *is a bounded subset such that its boundary is piecewise* C^2 *and locally convex at the corners, then the weak solution of the above problem is in* $H^2(\Omega)$, *i.e. the above mentioned operator* $\widehat{S}$ *has domain of definition* $H^2(\Omega) \cap H_0^1(\Omega)$.

If not noted otherwise we will have the assumptions of this theorem on Ω from now on. Most of the machinery would work, as weak solution makes sense for all uniformly positive operators, but using this theorem the weak form of the elliptic equation above can be written

$$\int_\Omega -\Delta u \cdot v = \int_\Omega fv \quad \forall v \in H_0^1(\Omega),$$

where the differential Δ operator is meant in the distributional sense.

1.3 Iteration methods

1.3.1 Conjugate gradient method

The contents of this section can be found as whole or as part in [6, 19, 43]. When treating partial differential equations numerically we often have to solve large scale linear problems in the classical form

$$Ax = b,$$

where A is some matrix and b is a given vector. There are mainly two types of methods solving a linear system like this: direct ones and iterative ones. Direct methods give exact solution to the equation. Most of them are associated to a matrix decomposition, LU decomposition, Cholesky decomposition, QR decomposition (with Gram-Schmidt orthogonalization or using Householder matrices) etc. Detailed description and analysis can be in most muerical mathematics book, for example in [19]. From a computational point of view direct methods has a drawback when solving large sparse matrix equations. The first and most stressful problem is that the given matrix is modified as the solution process advance, so filling-in the nonzero elements can cause serious storage problems. Although there are methods reducing fill-in e.g. incomplete LU decomposition or reordering (using e.g. Cuthill-Mckee or reverse Cuthill-McKee algorithms) the matrix when applying Cholesky decompostion to symmetric positive definite (SPD) matrices. Second problem is their sensitivity to ill-conditionness. For ill-behaved matrices with large condition number (these occur most prominently when discretizing partial differential equations) the numerical roundoff errors can be overwhelming. The most cited academic example is the Hilbert matrix. Since the direct methods are finite-step algorithms if they go on a sidetrack there is no way repairing. Thirdly some of these methods doesn't emploit special structures of the matrices so from a computational complexity view sometimes they are considered slow.

On the other hand with iteration methods, the matrix A is usually not modified at all, although for some methods we have to store some data to have the method running, but this need of storage is usually much smaller that those of the direct methods. There is also one more practical aspect that is in favor of iteration methods, that is when dealing with discretized partial differential equations we only have approximations of the continuous equations (often an order of the approximation is given). Hence there is no need for computing the exact solution to the discretized equation, it is enough to derive an approximate solution that is apporximately as close to the exact solution of the discretized equations as much approximation error we have made during discretization.

Krylov subspace methods

In the 1950's there was a major breakthrough in the theory of iteration methods, most prominently the introduction of the Lanczos method [31], the conjugate gradient method by Hestenes and Stiefel [22]. Properties of these methods and as well as their generalizations were analysed since then. These iterations fall into the wide class of Krylov subspace methods. In these methods as usual we have an initial guess x_0 to the solution of $Ax = b$ and form the residual vector $r_0 = Ax_0 - b$. The nth approximation

is calculated via a projection to the so-called Krylov subspace

$$\mathcal{K}_A = \left\{b, Ab, \ldots, A^{n-1}b\right\}.$$

Thus the for nth residual we have that for some polynomial q_n of degree $n-1$

$$r_n = q_n(A)r_0.$$

This is the starting point of most analysis of the properties of the methods.

Methods differ on the choice of the projection. The projections usually are chosen either minimize some norm of the projection error or minimize the norm of the residuals themselves. For SPD matrices, and in general uniformly positive bounded operators the choice of minimizing the projection error leads us to the conjugate gradient method.

Properties of the conjugate gradient method

Definition 1.21 (Conjugate gradient method (CG))**.**

Let $A \in B(H)$ be a uniformly positive operator, and $b \in H$ be arbitrary. The algorithm is as follows:

Initialize: $x_0 \in H$ arbitrary, $r_0 := Ax_0 - b$, $p_0 := r_0$,

while $r_n \neq 0$ or $r_n > \varepsilon$ **do**

$$\alpha_n := \frac{\langle r_n, p_n\rangle}{\langle Ap_n, p_n\rangle},\ x_{n+1} := x_n - \alpha_n p_n,\ r_{n+1} = r_n - \alpha p_n,$$
$$\beta_n := \frac{\langle Ar_{n+1}, p_n\rangle}{\langle Ap_n, p_n\rangle},\ p_{n+1} := r_{n+1} - \beta_n p_n.$$

end while

Proposition 1.22.

- *The above iteration converges to the solution of the equation $Ax = b$. If $\dim H < \infty$ the algorithm stops after finite steps, because of the A-orthogonality of the vectors p_k.*
- *Let $e_n = x_n - x$, and $\mathbb{P}_n^1 = \{p \in \mathbb{P} : \deg p \leq n, p(0) = 1\}$*

$$\frac{\|e_n\|_A}{\|e_0\|_A} \leq 2 \cdot \min_{p \in \mathbb{P}_n^1} \max_{\lambda \in \sigma(A)} |p(\lambda)|. \tag{1.2}$$

- *In the general case, using the extremal property of the Chebyshev polynomials, the right hand side yields the upper estimate $\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n$, where the $\kappa = \|A\|\|A^{-1}\|$ is the condition number of the operator A.*

For special classes of operators better convergence estimations can be obtained. First superlinear convergence result was first proved in 1954 by Hayes[21]. Superlin-

ear convergence replaces linear convergence in the sense that the linear convergence estimate of the form q^n with some $0 < q < 1$ can be replaced by an estimate q_n^n with $q_n \searrow 0$. For operators that are perturbations of the identity, i.e. $A = I + C$ where C compact this superlinear estimate can be obtained see e.g. [49],[35],[9]. We sketch the work of [49].

For simplicity we assume that C is positive and its eigenvalues λ_i are in non-increasing order. For this C a specially selected pair of polynomials Q_n, P_{n-1} where $P_{n-1} \in \mathbb{P}_n^1$ such that

$$Q_n(x) = 1 - (1+x)P_{n-1} \quad \text{and} \quad Q_n(\lambda_j) = 0 \quad \text{for} \quad j = 1, \ldots, n.$$

Hence using the minimizing property we have that

$$\|r_n\| \leq \|Q_n(C)r_0\|.$$

Here by choice of Q_n the action of $Q_n(C)$ in the direction to the eigenvectors correponding to the first n eigenvalues of C is cancelled, this insight is the reason behind the existence of a better (superlinear) convergence estimate. By precise calculations we obtain

$$\frac{\|r_n\|}{\|r_0\|} \leq \prod_{j=1}^{n} \frac{2\lambda_j}{1+\lambda_j} \leq \left(2 \frac{\sum_{j=1}^{n} \lambda_j}{n} \right)^n. \tag{1.3}$$

We may observe that the compactness implies that in the right hand side the expression in the brackets tends to zero. Thus we have superlinear convergence.

If C is of Hilbert-Schmidt class then in (1.3) and the inequality between the arithmetic and quadratic mean gives

$$\frac{\|r_n\|}{\|r_0\|} \leq \left(\frac{2\,|||C|||}{n} \right)^{n/2}.$$

In [9] using a different choice of polynomials they give a similar estimate. The choice here is that for $m > n = 2k$ we choose the polynomial P_n such that it vanishes in $\lambda_1, \ldots, \lambda_k$ and $\lambda_{m-k+1}, \ldots, \lambda_m$. Thus this polynomial is flat on both sides of the spectrum of $A = I + C$, hence the minmax expression in (1.2) can be efficiently estimated. Precise calculation shows that

$$\frac{\|e_n\|_A}{\|e_0\|_A} \leq \left(\frac{3\,|||C|||}{2n} \right)^{n/2}, \quad \text{for even} \quad n.$$

These two convergence estimates are asymptotically equivalent since

$$1 \leq \frac{\|r_n\|}{\|e_n\|_A} = \frac{\|Ae_n\|}{\|A^{1/2}e_n\|} \leq \|A^{1/2}\|.$$

although the second one gives better coefficients. Since in our work we give only asymptotics, from our point of view these to are essentially equivalent.

If we have a regular operator $A \in B(H)$ then

$$Ax = b \quad \Longleftrightarrow A^*Ax = A^*b.$$

Since A^*A is uniformly positive selfadjoint the solution of the original could be obtained applying the CG method to the second equation, this gives the Conjugate gradient for normal equations (CGN) method. A simple observation shows that we do not need to calculate the operator A^*A.

Definition 1.23 (Conjugate gradient method for normal equations (CGN)).

Let $A \in B(H)$ be a regular operator, and $b \in H$ be arbitrary. The algorithm is as follows:

Initialize: $x_0 \in H$ arbitrary, $s_0 := d_0 := A^*r_0$,

while $r_n \neq 0$ or $r_n > \varepsilon$ **do**

$$z_n = Ad_n,$$
$$\alpha_n = \frac{\langle r_n, z_n \rangle}{\|z_n\|^2}, \quad x_{n+1} = x_k - \alpha_n d_k, \quad r_{n+1} = r_n - \alpha_n z_n,$$
$$s_{n+1} = A^*r_{n+1},$$
$$\beta_k = \frac{\|s_{n+1}\|^2}{\|s_n\|^2}, \quad d_{n+1} = s_{n+1} + \beta_n d_n.$$

end while

Remark 1.24. If for the operator A we have $\inf\{\|Ax\| : \|x\| = 1\} \geq m > 0$ then the above algorithm converges.

The convergence of the CG method, as for almost all iteration methods, can be improved using some kind of preconditioning. A simple addtitive preconditioning idea is based on the following:

$$Ax = b \Longleftrightarrow (P + Q)x = b \Longleftrightarrow (I + P^{-1}Q)x = P^{-1}b,$$

and then the iteration is applied to the last equation. Classical examples of this preconditioning are the Jacobi and Gauss-Seidel iterations, which are preconditioned verisions

of the Richardson iteration.

One presentation of preconditioning is to achieve the goal to transform the eigenvalues of the operator close to 1, hence in (1.2) for a suitable chosen polynomial we may achieve a better estimation. Also if dealing with matrix equations and both P and Q are somewhat ill-behaved then these two problems may be dealt with seperately.

If P is a uniformly positive bounded operator, then essentially we are to an equation on the energy space H_P. A similar idea will appear in our proposal for preconditioning discretizations of elliptic problems, we will discuss this later on.

1.3.2 Newton method

The Newton method and its variants are widely used when solving nonlinear equations. The classical Newton method is the following.

Let $f : X \to X$ be a differentiable function on a Banach space, and we have to solve the nonlinear equation

$$f(x) = b, \text{ with some } b \in X.$$

The algorithm is as follows:

Initialize: $u_0 \in X$, $r_0 = b - F(x_0)$

while $\|r_n\| > \varepsilon$ **do**

$r_n = b - f(x_n)$,

$p_n = F'(u_n)^{-1} r_n$

update $x_{n+1} = x_n + p_n$

end while

The precise treatise of the convergence properties of the Newton method began with Kantorovich in the 1940's. Usually there is only local convergence, i.e. the initial guess x_0 has to be sufficiently close to the solution if we want to ensure convergence.

Sometimes convergence fails because the correction vector p_n is too long, thus instead of using p_n we use $\tau_n p_n$ usually with $0 < \tau_n < 1$. A well-chosen τ_n, called a damping parameter, may ensure convergence.

As the subproblem, solving the equation $f'(x_n)p_n = r_n$, is a linear one then as mentioned before iterative methods can be used. Thus it is useful to consider a variant where only an approximation of p_n is calculated.

We will use the following algorithm:

Theorem 1.25 (Damped inexact Newton method).
Let the function $F : X \to X$ be differentiable, and have the properties

(i) $\|F'(u)h\| \geq \lambda\|h\| \quad (u, b \in X)$ *with some constant* $\lambda > 0$ *independent of* u, h,

(ii) $\|F'(u) - F'(v)\| \leq L\|u - v\| \quad (u, v \in X)$ *with some constant* $L > 0$ *independent of* u, v.

The algorithm is as follows

Initialize: $u_0 \in X$, $r_0 = b - F(u_0)$

while $\|r_n\| > \varepsilon$ ***do***

$r_n = b - f(u_n)$,

find p_n *such that*

$$\|F'(u_n)p_n - r_n\| \leq \delta_n\|r_n\| \text{ with } 0 < \delta_n \leq \delta_0 < 1 \tag{1.4}$$

define $\tau_n = \min\left\{1, \frac{1-\delta_n}{(1+\delta_n)^2}\frac{\lambda^2}{L\|r_n\|}\right\}$,

update $x_{n+1} = u_n + \tau_n p_n$

end while

It has the following convergence properties:

- *The sequence* (u_n) *converges to the exact solution* u^* *of equation (2.16) as*

$$\|u_n - u^*\| \leq \|F(u_n) - b\| \to 0 \text{ monotonically.}$$

- *if* $\delta_n \equiv \delta_0$ *then we have linear convergence*

- *if* $\delta_n \leq const \cdot \|F(u_n) - b\|^\gamma$ $(0 < \gamma \leq 1)$ *then the convergence is locally of order* $1 + \gamma$, *that is the convergence is linear for* n_0 *steps, until* $\|F(u_n) - b\| \leq \varepsilon$, *where* ε *is at most* $(1 - \delta_n)\frac{1}{2L}$, *and further on* $\|u_n - u^*\| \leq Cq^{(1+\gamma)^{n-n_0}}$ *holds.*

Remark 1.26. The formula (1.4) indeed gives a bound for the error of the approximate solution of the linearized equation

$$F'(u_n)p_n = r_n.$$

Remark 1.27. Similar converge result holds if we relax the condition of Lipschitz continuity to a local Lipschitz continuity such as

$$\|F'(u) - F'(v)\| \leq L(r)\|u - v\| \quad (u, v \in X, \|u\| < r, \|v\| < r)$$

where the function $L : \mathbb{R}^+ \to \mathbb{R}^+$ is non-decreasing.

2. CONVERGENCE ESTIMATES USING THE HILBERT-SCHMIDT NORM

In this chapter we'll provide iterative methods and related mesh independent convergence estimates for the FEM solution of classes of linear and nonlinear elliptic systems. We prove superlinear convergence for both classes of equations.

In the first section we consider symmetric linear coupled PDE systems. In this case the proposed solution method is a preconditioned conjugate method (PCG). We show superlinear convergence giving an explicit convergence extimation. We achieve this that by preconditioning with the stiffness matrix we arrive to matrix equation equation and the Frobenius norm that shows up in the convergence estimate of the conjugate gradient method. We give mesh independent upper estimate of this Frobenius by the Hilbert-Schmidt norm of the n-tuple of inverse Laplacians defined on $H_0^1(\Omega)$.

In the nonlinear case we show that a preconditioned damped inexact Newton iteration has superlinear convergence, and since the linearization of the equation i.e. the Newton iteration's linear subproblem is of the form of the equations of the first section, the PCG applied to the subproblem has the same convergence property that is proved in the first section.

2.1 Linear symmetric elliptic systems

First we consider an abstract problem, with the superlinear mesh independent convergence property, proved in [27]. In the following section we consider an elliptic partial differential system and we prove that it has an abstract form that is discussed in Section 2.1.1. In the next section we state the main mesh independence results and finally we give a numerical solution method, actually finite element method, and we give our results on the testing of the proved theoretical results, with code written in Matlab.

2.1.1 An abstract problem

Let H be a separable Hilbert space and g an arbitrary element in H. Let us consider the linear equation

$$Bu = g, \tag{2.1}$$

with a linear operator B satisfying the following conditons:

1. (i) B has the form of $B = S + Q$, where S is densely defined (in our application unbounded), Q is bounded and both are linear self-adjoint operators on the Hilbert space H,
2. (ii) $\exists p > 0$: $\langle Su, u\rangle \geq p\|u\|^2$, $(u \in D(S))$,
3. (iii) $\langle Qu, u\rangle \geq 0$, $(u \in H)$,
4. (iv) the operator $S^{-1}Q$ defined on H_S is compact and of type Hilbert-Schmidt.

We replace (2.1) by its preconditioned form $(I+S^{-1}Q)u = S^{-1}g$ which is equivalent to the following weak formulation, as defined in Definition 1.4.

$$\langle u, v\rangle_S + \langle Qu, v\rangle = \langle g, v\rangle \, (\forall v \in H_S), \tag{2.2}$$

which has a unique solution $u \in H_S$ by conditions (ii) and (iii).

Now equation (2.2) is solved numerically using Galerkin discretization.
Let $V = \operatorname{span}\{\varphi_1, \ldots, \varphi_k\} \subset H_S$ be a given finite-dimensional subspace,

$$\mathbf{S} = \{\langle \varphi_i, \varphi_j\rangle_S\}_{i,j=1}^k \text{ and } \mathbf{Q} = \{\langle Q\varphi_i, \varphi_j\rangle\}_{i,j=1}^k$$

be the Gram matrices corresponding to S and Q. Seeking the solutions in V in the form of $u_V = \sum_{j=1}^k c_j\varphi_j$ we obtain a finite linear system

$$(\mathbf{S} + \mathbf{Q})c = \mathbf{b}, \tag{2.3}$$

with $\mathbf{c} = (c_1, \ldots, c_k)^T$ and $\mathbf{b} = \{\langle g, \varphi_j\rangle\}_{j=1}^k$. By the preceding assumptions the matrix $\mathbf{S} + \mathbf{Q}$ is symmetric positive definite.

As we did it at the abstract case, we use the matrix $\mathbf{S}$ as a preconditioner for system (2.3). After preconditioning we obtain the system

$$(\mathbf{I} + \mathbf{S}^{-1}\mathbf{Q})\mathbf{c} = \widetilde{\mathbf{b}}, \tag{2.4}$$

where $\widetilde{\mathbf{b}} = \mathbf{S}^{-1}\mathbf{b}$, and $\mathbf{I}$ denotes the identity matrix on $\mathbb{R}^k$. Now we apply the conjugate gradient method for the solution of system (2.4). The following theorems are proved in [27].

Theorem 2.1. *Let assumptions (i)-(iv) hold. Then*

$$|||\mathbf{S}^{-1}\mathbf{Q}||| \leq |||S^{-1}Q|||,$$

where $|||\cdot|||$ denotes the Frobenius norm of a matrix and the Hilbert-Schmidt norm of a compact operator respectively.

Proof. Let (λ_j) denote the eigenvalues of $\mathbf{S}^{-1}\mathbf{Q}$ in descending order with multiplicity. Since $\mathbf{S}^{-1}\mathbf{Q}$ is adjoint w.r.t. the $\langle\cdot,\cdot\rangle_S$ scalar product

$$\langle S^{-1}Qu, v\rangle_S = \langle Qu, v\rangle = \langle u, Qv\rangle = \langle S^{-1}u, Qv\rangle_S = \langle u, S^{-1}Qv\rangle_S$$

, it has orthonormal eigenvectors $(\mathbf{c_j})$ corresponding to the eigenvalues. And also by definition there are elements $(u_j) \subset V \subset H$ corresponding elements those vectors. This set of orthonormal vectors may be extended to a complete orthonormal system in H with elements $(u_{k+i}) \subset V^{\perp} \subset H \quad (i = 1, 2, \dots)$. By the definition of the Hilbert-Schmidt norm (Definition 1.11) we have that

$$|||\mathbf{S}^{-1}\mathbf{Q}||| = \sum_{j=1}^{k}\langle Qu_j, u_j\rangle \leq \sum_{j=1}^{\infty}\langle Qu_j, u_j\rangle = |||S^{-1}Q|||\,.$$

□

Corollary 2.2. *The conjugate gradient method applied to system (2.4) yields the following estimate:*

$$\frac{\|e_n\|}{\|e_o\|} \leq \left(\frac{3\,|||S^{-1}Q|||^2}{2n}\right)^{n/2},$$

if $n \in \mathbb{N}$ is even and $n \geq (3/2)\,|||S^{-1}Q|||^2$. This estimate is independent of the subspace V.

2.1.2 A class of symmetric elliptic systems

In this section we consider self-adjoint second order elliptic boundary value systems and their finite element discretizations. We prove that this problem has an abstract form that satisfies the assumptions in Section 2.1.1.

The linear elliptic system

Let $d \leq 3$ and $\Omega \subset \mathbb{R}^d$ be a bounded domain. We consider the elliptic problem

$$\begin{cases} -\mathrm{div}(G_1\nabla u) + d_{11}u + d_{12}v = g_1 \\ -\mathrm{div}(G_2\nabla v) + d_{21}u + d_{22}v = g_2 \\ u_{|\partial\Omega} = 0 \\ v_{|\partial\Omega} = 0 \end{cases} \tag{2.5}$$

under the following assumptions:

(a) $\partial\Omega$ is piecewise C^2 and Ω is locally convex at the corners,

(b) $G_1, G_2 \in C^1(\bar{\Omega}, \mathbb{R}^{d\times d})$, both symmetric at the points of $\bar{\Omega}$, and $\exists 0 < m \leq M < \infty$:

$$m\|\xi\|^2 \leq G_i(x)\xi \cdot \xi \leq M\|\xi\|^2, \quad (x \in \bar{\Omega},\ \xi \in \mathbb{R}^d,\ i = 1, 2),$$

(c) $D = D^T = \{d_{ij}\}_{i,j=1}^2 \in C(\bar{\Omega}, \mathbb{R}^{2\times 2})$ and $D \geq 0$ on $\bar{\Omega}$,

(d) $g_i \in L^2(\Omega)$ $(i = 1, 2)$.

Remark 2.3. The number of equations in (2.5) can be any positive integer n, for simplicity we have chosen $n = 2$.

In the next two sections we shall prove that the system (2.5) corresponds to a operator equation of the form we mentioned in Section 2.1.1.

Let H be the product space $L^2(\Omega) \times L^2(\Omega)$ with the inner-product

$$\left\langle \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right\rangle = \langle u_1, v_1\rangle_{L^2(\Omega)} + \langle u_2, v_2\rangle_{L^2(\Omega)}.$$

The notation $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ will hold from now on. The operators S and Q are defined as

$$S\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} -\mathrm{div}(G_1\nabla u_1) \\ -\mathrm{div}(G_2\nabla u_2) \end{pmatrix}, \quad (\mathbf{u} \in D(S)), \tag{2.6}$$

and

$$Q\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = D \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}, \quad \left(\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \in H\right), \tag{2.7}$$

where $D(S) = (H^2(\Omega) \cap H_0^1(\Omega)) \times (H^2(\Omega) \cap H_0^1(\Omega))$. Therefore the system can be written as

$$(S + Q)\mathbf{u} = \mathbf{g}, \tag{2.8}$$

which has the form of (2.1).

Proving assumptions

The following easy computations show that the assumptions of Section 2.1.1 hold for the operator equation (2.8).

By the Green formula and the homogeneous Dirichlet boundary conditions

$$\left\langle S\begin{pmatrix}u\\v\end{pmatrix},\begin{pmatrix}u\\v\end{pmatrix}\right\rangle = -\int_\Omega (\operatorname{div}(G_1\nabla u)\cdot\nabla u + \operatorname{div}(G_2\nabla v)\cdot\nabla v) = \int_\Omega (G_1\nabla u\cdot\nabla u + G_2\nabla v\cdot\nabla v),$$

hence with assumption (b) we have

$$m\int_\Omega (|\nabla u|^2+|\nabla v|^2) \le \left\langle S\begin{pmatrix}u\\v\end{pmatrix},\begin{pmatrix}u\\v\end{pmatrix}\right\rangle \le M\int_\Omega (|\nabla u|^2+|\nabla v|^2). \tag{2.9}$$

The Friedrichs-Poincaré inequality gives

$$\int_\Omega (|\nabla u|^2+|\nabla v|^2) \ge \nu\int_\Omega (u^2+v^2),\ \left(\begin{pmatrix}u\\v\end{pmatrix}\in D(S)\right),$$

so we have that assumption (ii) is fulfilled with $p=\nu m$, and beyond this by (2.9) we also have that the energy space H_S coincides with the space $(H_0^1(\Omega))\times(H_0^1(\Omega))$.

It is obvious that assumption (iii) is fulfilled.

Since S is the pair of the symmetric operators $S_1, S_2 : L^2(\Omega)\to L^2(\Omega)$, which are symmetric and superjective by assumptions (a) and (b), therefore self-adjoints by Theorem 1.5 and Theorem 1.20, hence their pair S is also self-adjoint. It is obvious that $D\in L^\infty(\Omega)$ in the sense that $d_\infty = \sup_{\bar\Omega}\|D(x)\|_2 < \infty$. Hence assumption (i) also holds.

Let us take $\mathbf{u}\in H$ and $\mathbf{v}\in H_S$.

$$\left|\left\langle S^{-1}Q\begin{pmatrix}u_1\\u_2\end{pmatrix},\begin{pmatrix}v_1\\v_2\end{pmatrix}\right\rangle_S\right| = \left|\int_\Omega D\mathbf{u}\cdot\mathbf{v}\right| \le$$
$$d_\infty\int_\Omega \sqrt{(\|u_1\|^2+\|u_2\|^2)}\sqrt{(\|v_1\|^2+\|v_2\|^2)} \le$$
$$d_\infty\sqrt{\int_\Omega(\|u_1\|^2+\|u_2\|^2)}\sqrt{\int_\Omega(\|v_1\|^2+\|v_2\|^2)} =$$
$$d_\infty\left\|\begin{pmatrix}u_1\\u_2\end{pmatrix}\right\|_H\left\|\begin{pmatrix}v_1\\v_2\end{pmatrix}\right\|_H \le C\left\|\begin{pmatrix}u_1\\u_2\end{pmatrix}\right\|_H\left\|\begin{pmatrix}v_1\\v_2\end{pmatrix}\right\|_S,$$

where $C=\frac{d_\infty}{\sqrt{m\nu}}$. From this we conclude that the operator $S^{-1}Q$ is bounded from H to H_S.

Since the embedding $H_S\hookrightarrow H$ is compact, because the embedding of the Sobolev space $H_0^1(\Omega)$ to $L^2(\Omega)$ is compact, therefore the operator $S^{-1}Q : H_S\to H_S$ is also compact.

We remark that the exsistence and uniqueness of the weak solution of (2.5) can be easily verified on the usually way as in elliptic problems even with weaker assumptions

to Ω or the matrices G_i, D.

So far we have proved all assumptions (i)-(iv), all but the Hilbert-Schmidt property of $S^{-1}Q$.

The Hilbert-Schmidt property

For the requisite estimate of the norm $|||S^{-1}Q|||$, we use the variational property of the eigenvalues of a compact self-adjoint operator, namely the min-max theorem Proposition 1.10 gives an estimate for the eigenvalues of $S^{-1}Q$. From now on λ_j denotes the jth eigenvalue of the operator $S^{-1}Q$. By equation (2.9) we have

$$\frac{\langle S^{-1}Q\mathbf{u}, \mathbf{u}\rangle_S}{\langle \mathbf{u}, \mathbf{u}\rangle_S} = \frac{\int_\Omega D\mathbf{u}\cdot\mathbf{u}}{\int_\Omega S\mathbf{u}\cdot\mathbf{u}} \leq \frac{d_\infty}{m}\frac{\int_\Omega \mathbf{u}\cdot\mathbf{u}}{\int_\Omega |\nabla u_1|^2 + |\nabla u_2|^2}, \tag{2.10}$$

where d_∞ is as defined before, so we have

$$\lambda_{j+1} = \inf_{\mathbf{u}_1,\dots\mathbf{u}_j} \sup_{\mathbf{u}\perp\mathbf{u}_1,\dots,\mathbf{u}_j} \frac{\langle S^{-1}Q\mathbf{u}, \mathbf{u}\rangle_S}{\langle \mathbf{u}, \mathbf{u}\rangle_S} \leq \frac{d_\infty}{m}\lambda'_{j+1},$$

where

$$\lambda'_{j+1} = \inf_{\mathbf{u}_1,\dots\mathbf{u}_j} \sup_{\mathbf{u}\perp\mathbf{u}_1,\dots,\mathbf{u}_j} \frac{\int_\Omega \mathbf{u}\cdot\mathbf{u}}{\int_\Omega |\nabla u_1|^2 + |\nabla u_2|^2}.$$

Now we define the operators S' and Q' in the same way as we did it in Section 2.1.2, with $G'_1 \equiv G'_2 \equiv I$ and $D' \equiv I$, where I is the identity matrix of $\mathbb{R}^2$, defining the system

$$\begin{cases} -\mathrm{div}(\nabla u) + u = g_1 \\ -\mathrm{div}(\nabla v) + v = g_2 \\ u_{|\partial\Omega} = 0 \\ v_{|\partial\Omega} = 0. \end{cases} \tag{2.11}$$

We observe that by Proposition 1.10 λ'_{j+1} is the $(j+1)$th eigenvalue of $S'^{-1}Q'$. Since this system is decoupled, the operator $S'^{-1}Q'$ is the pair of the two operators $(-\Delta)^{-1}$, hence the eigenvalues of $S'^{-1}Q'$ are the same as the eigenvalues of $(-\Delta)^{-1}$, with doubled multiplicity.

A consequence of Proposition 1.10 is that if $\Omega \subset \Omega'$ then the eigenvalues μ_j of $(-\Delta)^{-1}$ has the following property (see in [1]): $\mu_j(\Omega) \leq \mu_j(\Omega')$. Now take $\Omega' = R$ the rectangle that contains a translate of Ω. Since the eigenvalues of $(-\Delta)^{-1}$ are known on a rectangle (see in [1]) we may summarize the above results in the following theorem:

Theorem 2.4. *With the previous notations, the operator $S^{-1}Q$ is Hilbert-Schmidt if and only if $N \leq 3$, and the following computable estimates are true for its Hilbert-Schmidt norm depending on space dimension:*

· $d = 1$, *take* $R = [0, a_1]$, *then*

$$\sigma_1^2 = \left|\left|\left| S^{-1}Q \right|\right|\right|^2 \leq \frac{4d_\infty^2}{m^2\pi^2} \sum_{i=1}^{\infty} \left(\frac{i^2}{a_1^2}\right)^{-2}$$

· $d = 2$, *take* $R = [0, a_1] \times [0, a_2]$, *then*

$$\sigma_2^2 = \left|\left|\left| S^{-1}Q \right|\right|\right|^2 \leq \frac{4d_\infty^2}{m^2\pi^2} \sum_{i,j=1}^{\infty} \left(\frac{i^2}{a_1^2} + \frac{j^2}{a_2^2}\right)^{-2}$$

· $d = 3$, *take* $R = [0, a_1] \times [0, a_2] \times [0, a_3]$, *then*

$$\sigma_3^2 = \left|\left|\left| S^{-1}Q \right|\right|\right|^2 \leq \frac{4d_\infty^2}{m^2\pi^2} \sum_{i,j,k=1}^{\infty} \left(\frac{i^2}{a_1^2} + \frac{j^2}{a_2^2} + \frac{k^2}{a_3^2}\right)^{-2}.$$

Remark 2.5.

· The assumptions in (2.5) can be more general as mentioned in [27].

· As mentioned before the number of the equations in the system (2.5) may be any positive integer n. One may prove an analogous statement as in Theorem 2.4 replacing $4 = 2^2$ by 2^n.

2.1.3 *The mesh independence result for the finite element method*

The formulation of the finite element method and the solution algorithm

Let $V_h^0 \subset H_0^1$ be a finite element subspace and define $V_h = V_h^0 \times V_h^0 \subset H_S$. We look for the approximate solution of (2.5) in V_h. That is we look for a $\mathbf{u_h} \in V_h$ such that for all $\mathbf{w_h} \in V_h$

$$\int_\Omega \left(G_1 \nabla u_{h,1} \cdot \nabla w_{h,1} + G_1 \nabla u_{h,1} \cdot \nabla w_{h,1} + D\mathbf{u} \cdot \mathbf{u}\right) = \int_\Omega \left(g_1 w_{h,1} + g_2 w_{h,2}\right),$$

which is equivalent to the following linear algebraic system

$$(G_h + D_h)c = g_h, \tag{2.12}$$

where the stiffness and mass matrices G_h and D_h are defined as usual. We solve it by preconditioning with G_h. The algorithm is as follows

Initialize: $c_0 \in \mathbb{R}^k$, $r_0 = c_0 + y_0 - g_h$ where y_0 solves $G_h y_0 = D_h c_0$, $p_0 = r_0$

while $\|r_n\| > \varepsilon$ **do**

calculate

$$\alpha_n = \frac{\langle G_h r_n, p_n \rangle}{\langle (G_h + Q_h) p_n, p_n \rangle},$$

update $c_{n+1} = c_n - \alpha_n p_n$,

update $r_{n+1} = r_n - \alpha_n (p_n - y_n)$, where $G_h y_n = D_h p_n$,

calculate

$$\beta_n = \frac{\langle (G_h + Q_h) p_n, r_{n+1} \rangle}{\langle (G_h + Q_h) p_n, p_n \rangle},$$

update $p_{n+1} = r_{n+1} - \beta_n p_n$.

end while

The advantage of this method that in some cases there are fast solvers for the preconditioner equations above (see eg. [12, 42]). Now we can state our main result:

Theorem 2.6. *The above described preconditioned conjugate gradient method applied to (2.12) has the following mesh independent estimate of its error:*

$$\frac{\|e_n\|}{\|e_0\|} \leq \left(\frac{3}{2n} \sigma_d^2 \right)^{n/2},$$

where σ_d is defined in Theorem 2.4 and $n \geq (3/2)\sigma_d^2$.

2.1.4 Numerical experiments

Here we illustrate the preceding theoretical results by some numerical tests. The code was written in Matlab. Even though the discretizations are the simplest, the stiffness and mass matrices were built and not generated by the tools of Matlab. We also solved the linear equations with the solver of Matlab, the error was computed as the S-norm of the difference of the iteration solution and the solution given us by Matlab.

The program is on the unit square $[0,1] \times [0,1]$, with $G_i \equiv I$ and D is some matrix satisfying assumption (b) of Section 2.1.2. Equidistant mesh and the canonical Courant-elements were used, in the tabs m denotes the number of intervals in $[0,1]$, so finite element mesh consists of $2m^2$ triangles. In the following tabs σ_d^2 is calculated by Theorem 2.4 and

$$\sigma_d^{*2} = \sup_{n \geq (3/2)\sigma_d^2} \{ (\|e_n\|_S / \|e_0\|_S)^{2/n} (2n/3) \| \}$$

is the constant for the superlinear convergence. We used two different functions g_1, g_2 on the right side of 2.8. The results are:

	10	20	30	40	50	60
σ_d^2	4.7619	5.2360	5.3852	5.4571	5.4994	5.5271
σ_d^{*2}	2.7016	2.6072	2.4889	2.5957	2.5275	2.4720

	10	20	30	40	50	60
σ_d^2	2.1885	2.3046	2.3441	2.3640	2.3760	2.3840
σ_d^{*2}	2.8837	2.6180	2.4828	2.3916	2.3229	2.2682

2.2 Semilinear elliptic systems

In this section we propose we propose an inner-outer (damped inexact Newton plus PCG) iteration for the finite element discretization of a class of nonlinear elliptic systems. Our aim is to show mesh independent superlinear convergence of the overall iteration. The linearized equations will be solved by a preconditioned conjugate gradient method. It is known that the Newton method has quadratic convergence when the exact solution of the linearized equation is given. Instead of this, one may solve the linearized equation in an inexact way, mainly with applying an iteration method, in this paper we consider a preconditioned conjugate gradient method. This way we lose the quadratic convergence of the outer Newton iterations, but we may ensure superlinear convergence as we control the inaccuracy of the inner iteration.

2.2.1 The PDE system

We consider the class of semilinear PDE-systems described below, which has the short form

$$\begin{cases} -\Delta \underline{u}(x) + f(x, \underline{u}(x)) = \underline{g}(x) \\ \underline{u}|_{\partial\Omega} = \underline{0}, \end{cases} \tag{2.13}$$

where $\underline{u} = (u_1, u_2, \ldots, u_s)^T, \underline{g} = (g_1, g_2, \ldots, g_s)^T$. In this work all operators, like $\Delta, \nabla, |_{\partial\Omega}$, are meant coordinatewise.
We impose the assumptions

[P1] $\partial\Omega \subset \mathbb{R}^d$ ($d = 2$ or 3) is piecewise C^2 and Ω is locally convex at the corners,

[P2] $g_i \in L^2(\Omega)$ $(i = 1, 2, \ldots, s)$ on Ω,

[P3] $f : \Omega \times \mathbb{R}^s \to \mathbb{R}^s$, for a.e. $x \in \Omega$ $f(x, \xi)$ has a potential $\psi : \Omega \times \mathbb{R}^s \to \mathbb{R}$, i.e. $f = \partial_\xi \psi$ and is differentiable w.r.t. ξ, and in these points the Jacobians are symmetric positive semidefinite,

[P4] for a.e. $x \in \Omega$ the Jacobians $\partial_\xi f(x,\xi)$ are uniformly bounded in ξ by a symmetric matrix $M(x)$, where the eigenvalues $\mu_j(x)$ of $M(x)$ are bounded $0 \le \mu_j(x) \le c_1$, with some constant $c_1 > 0$,

[P4'] the eigenvalues $\lambda_j^{(f)}(x,\xi)$ $(j = 1,\dots,s)$ of the Jacobians $\partial_{\underline{\xi}} f(x,\xi)$ are bounded as follows

$$0 \le \lambda_j^{(f)}(x,\xi) \le c_2 + c_3 \sum_{j=1}^{s} |\xi|^{p-2},$$

for some constants $c_2, c_3 > 0$ and $p \ge 2$,

[P5] the derivative of f is Lipschitz continuous, that is there exists a constant C that $\|\partial_\xi f(x,\xi_1) - \partial_\xi f(x,\xi_2)\|_2 \le C\|\xi_1 - \xi_2\|_2$ for a.e. $x \in \Omega$,

[P5'] the derivative of f is locally Lipshcitz continuous, that is there exists a function $C : (0,\infty) \to (0,\infty)$ that $\|\partial_\xi f(x,\xi_1) - \partial_\xi f(x,\xi_2)\|_2 \le C(r)\|\xi_1 - \xi_2\|_2$ for a.e. $x \in \Omega$ if $\|\xi_1\|, \|\xi_2\| \le r$.

Definition 2.7. Let $H_1,\dots,H_k$ be Hilbert spaces. Then $H = H_1 \times \cdots \times H_k$ equipped with the inner product

$$\langle \underline{u}, \underline{v} \rangle_H = \sum_{i=1}^{k} \langle u_i, v_i \rangle_{H_i}$$

is a Hilbert space, with the notation $\underline{u} = (u_1,\dots,u_k) \in H$.

The weak formulation of this equation is that we seek the solution $\underline{u} \in \mathcal{H} = (\mathcal{H}, \langle\cdot,\cdot,\rangle) = (H_0^1(\Omega))^s$ that satisfies for all $\underline{v} \in (H_0^1(\Omega))^s$

$$\int_\Omega (\nabla \underline{u} \cdot \nabla \underline{v} + f(x,\underline{u}) \cdot \underline{v}) = \int_\Omega \underline{g} \cdot \underline{v}. \tag{2.14}$$

Here the operator ∇ is also meant coordinatewise.

Remark 2.8. $\mathcal{H} = (H_0^1(\Omega))^s$ coincides with the energy space of the unbounded operator $S : D(S) \subset (L^2(\Omega))^s \to (L^2(\Omega))^s$, the coordinatewise Laplacian. That is $\mathcal{H}$ is the completition of the space $(D(S), \langle\cdot,\cdot\rangle_S)$ where $\langle \underline{u}, \underline{v} \rangle_S = \int_\Omega S\underline{u} \cdot \underline{v} = \int_\Omega \nabla \underline{u} \cdot \nabla \underline{v}$ is the energy scalar product.

In the following $\langle\cdot,\cdot\rangle$ will always denote the above mentioned scalar product, and $\|\cdot\|$ will denote the induced norm.

The weak formulation of this equation also has its (equivalent) variational form, that is we seek the solution $\underline{u} \in \mathcal{H}$ that minimizes the function $\phi : \mathcal{H} \to \mathbb{R}$

$$\phi(\underline{u}) := \int_\Omega \left(\frac{1}{2} |\nabla \underline{u}|^2 + \psi(x,\underline{u}) - \underline{g} \cdot \underline{u} \right). \tag{2.15}$$

By assumption [P3] ψ is a convex function, and therefore ϕ is also convex. By (2.15) we have that ϕ is also coercive, therefore the function ϕ has a unique minimum (see e.g. [51]), and hence equation (2.14) has a unique weak solution.

2.2.2 Abstract form

Equation (2.14) may be considered as an equation on the space $\mathcal{H}$

$$F(\underline{u}) = \underline{b}, \tag{2.16}$$

where $F(\underline{u})$ and $\underline{b}$ are the Riesz representation vectors defined by the left and right-hand sides of (2.14) respectively, thus it is meant by the weak formula

$$\langle F(\underline{u}), \underline{v} \rangle = \langle \underline{b}, \underline{v} \rangle.$$

Proposition 2.9. *From assumptions [P1-P5] we have that*

(1) $F : \mathcal{H} :\to \mathcal{H}$ *is differentialbe in the Gateaux sense;*

(2) F *is regular, and* $\|F'(u)h\| \geq \|h\|$ *independent of* u, h*;*

(3) F *has the form* $F = I + N$*, where* I *is the identity operator on* $\mathcal{H}$*,* N *is also differentiable and for all* $u \in \mathcal{H}$*,* $N'(u)$ *is a compact self-adjoint operator, further it is Hilbert-Schmidt;*

(4) the operators $N'(u)$ *are uniformly majorized, that is there exists a compact positive self-adjoint Hilbert-Schmidt operator such that for all* $u \in \mathcal{H}$*,* $N'(u) \leq K$ *in the sense* $\langle N'(u)h, h \rangle \leq \langle Kh, h \rangle$*,* $\forall h \in \mathcal{H}$*;*

(5) if we have [P4'] instead of [P4] we only have the operators $N'(u)$ *are locally uniformly majorized, that is for all* $r > 0$ *there exists a compact positive self-adjoint Hilbert-Schmidt operator such that* $N'(u) \leq K(r)$ *in the sense* $\langle N'(u)h, h \rangle \leq \langle K(r)h, h \rangle$*,* $\forall h \in \mathcal{H}$*, for all* $\|u\| \leq r$*;*

(6) N' *is Lipschitz continous with Lipschitz constant* L*;*

(7) if [P5'] holds only instead of [P5] then N' *is only locally Lipschitz continous, with the function* $L : (0, \infty) \to (0, \infty)$*.*

Proof. Using assumptions [P3-P4] and the law of derivation under integral sign we have that the F is indeed differentiable and its weak form is

$$\langle F'(\underline{z})\underline{u}, \underline{v} \rangle = \int_\Omega \left(\nabla \underline{u} \cdot \nabla \underline{v} + f_{\underline{\xi}}(x, \underline{z})\underline{u} \cdot \underline{v} \right).$$

This gives (1) and the first part of (3), the rest of (3) comes from the exact same reasoning why the operator $S^{-1}Q$ was Hilbert-Schmidt in the previous section when we dealt with linear equations.

The positivity of the Jacobians $f_{\underline{\xi}}(x, \underline{u})$ give 2.

Assumption [P4] gives that $N'(u) \leq c_1 \cdot S^{-1}$, where S^{-1} is Hilbert-Schmidt according to the previous section, this gives (4). As for (5) we may write

$$\langle N'(\underline{z})\underline{v}, \underline{v}\rangle = \int_\Omega f_{\underline{\xi}}(x, \underline{z})\underline{v} \cdot \underline{v} \leq \int_\Omega c_2 \underline{v} \cdot \underline{v} + \int_\Omega c_3 |\underline{z}|^{p-2}\underline{v} \cdot \underline{v}.$$

Here only the second expression is interesting, the first one can be dealt with the same way as we did when we proved (4) before.

$$\int_\Omega |\underline{z}|^{p-2}\underline{v}\cdot\underline{v} \leq \||\underline{z}|\|_p \sum_i \|v_i\|_p \|v_i\|_p \leq \|z\|_p \cdot \|\underline{v}\|_{\mathcal{H}} \leq \||\underline{z}|\|_p \sum_i \|v_i\|_p \|v_i\|_p \leq C\|\underline{z}\|_{\mathcal{H}} \cdot \|\underline{v}\|_{\mathcal{H}}.$$

Here we used the generalized Hölder inequality for the exponents $\frac{p-2}{p} + \frac{1}{p} + \frac{1}{p} = 1$ and the Sobolev embedding theorem respectively. This gives the locally uniform bound in (5).

Proof of 6 is similar to (7) with a little less pain, so we omit it for brevity. As for (7) we conclude as follows. Let $\underline{z}, \underline{w} \in \mathcal{H}$ arbitrary vectors, $C(r)$ as int [P5'] and $\underline{u}, \underline{v} \in \mathcal{H}$ such that $\|\underline{u}\|, \|\underline{v}\| \leq r$. We may write

$$|\langle (N'(\underline{u}) - N'(\underline{v}))\,\underline{z}, \underline{w}\rangle| \leq \int_\Omega C(r)|\underline{u} - \underline{v}|\underline{z} \cdot \underline{w}.$$

This time we use the Hölder inequality for the exponents $\frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$ and the Sobolev embedding theorem respectively, thus we arrive to the estimation

$$|\langle (N'(\underline{u}) - N'(\underline{v}))\,\underline{z}, \underline{w}\rangle| \leq L(r)\|\underline{u} - \underline{v}\|\|\underline{z}\|\|\underline{w}\|,$$

where $L(r) = C(r) \cdot C^3$ where C is the norm of the embedding $H_0^1(\Omega) \hookrightarrow L^3(\Omega)$. Thus taking supremum over $\{\underline{z}, \underline{w} : \|\underline{z}\| = \|\underline{w}\| = 1\}$ we arrive to (7). □

Remark 2.10. From (1) and (2) of Proposition 2.9 and using a Hadamard type theorem [39],[51] we have an another proof that (2.14) has a unique solution.

Now we may state a convergence theorem of the DIN method.

Theorem 2.11. *As we can see the function F defined above satisfies the conditions of the DIN method if Theorem* 1.25. *Hence it converges to the solution from an arbitrary initial vector $\underline{u}_0$. And by carefully choosing the damping parameters we may obtain superlinear convergence.*

2.2.3 FEM discretization

Discretization of the PDE system

We consider the finite element discretization of the PDE system above. That is we have a finite element subspace $\mathcal{V}_h \subset \mathcal{H}$ with $\mathcal{V}_h = (V_h)^s = \text{span}\,(\underline{w}_h^i)_{i=1}^m$, where V_h is a finite element subspace in $H_0^1(\Omega)$. Then we seek the element $\underline{u}_h \in \mathcal{V}_h$ that satisfies

$$\int_\Omega (\nabla \underline{u}_h \cdot \nabla \underline{v}_h + f(x, \underline{u}_h) \cdot \underline{v}_h) = \int_\Omega \underline{g} \cdot \underline{v}_h, \text{ for all } \underline{v}_h \in \mathcal{V}_h. \tag{2.17}$$

This equation could also be understood as an equation on the Hilbert-space $\mathcal{V}_h$ (endowed with the inherited inner product $\langle \cdot, \cdot \rangle$)

$$F_h(\underline{u}) = \underline{b}_h, \tag{2.18}$$

where $F_h(\underline{u})$ and $\underline{b}_h$ are the Riesz representation vectors defined by the left and right-hand sides of (2.17) respectively.

Proposition 2.12. *$F_h(\underline{u})$ is the projection of $F(\underline{u})$ onto the subspace $\mathcal{V}_h$. It inherits all the analogous properties of F those are mentioned in Proposition 2.9.*

Corollary 2.13. *By Remark 2.10 we have that* (2.17) *also has a unique solution.*

Thus we are lead to the problem: find the coefficients $\mathbf{c} = (c_j)_{j=1}^m$ such that $\underline{u}_h = \sum c_j \underline{w}_h^j$ satisfies

$$\int_\Omega \left(\nabla \underline{u}_h \cdot \nabla \underline{w}_h^j + f(x, \underline{u}_h) \cdot \underline{w}_h^j\right) = \int_\Omega \underline{g} \cdot \underline{w}_h^j, \text{ for } j = 1, \ldots, m.$$

This gives rise to a nonlinear algebraic system of the following form, practically after preconditioning with $\mathbf{S_h}$ defined in the next section:

$$\mathbf{c} + N_h(\mathbf{c}) = \mathbf{b}. \tag{2.19}$$

Discretization of the linearized equation

From the above formulas we have that the linearization of F_h is

$$\langle F_h'(\underline{u})\underline{p}, \underline{v} \rangle = \int_\Omega \left(\nabla \underline{p} \cdot \nabla \underline{v} + \partial_{\underline{u}} f(x, \underline{u})\underline{p} \cdot \underline{v}\right) = \left\langle \underline{p}, \underline{v} + S^{-1} Q_{\underline{u}} \underline{v} \right\rangle, \text{ for all } \underline{p}, \underline{v} \in \mathcal{V}_h. \tag{2.20}$$

We may define the stiffness and mass matrices respectively as

$$\mathbf{S_h} = \left[\langle \underline{w}_h^i, \underline{w}_h^j \rangle\right]_{i,j=1}^m, \ \mathbf{D_h}(\underline{\mathbf{u}}) = \left[\langle S^{-1} Q_{\underline{u}} \underline{w}^i{}_h, \underline{w}^j{}_h \rangle\right]_{i,j=1}^m = \left[\langle Q_{\underline{u}} \underline{w}^i{}_h, \underline{w}^j{}_h \rangle_{(L_2)^s}\right]_{i,j=1}^m. \tag{2.21}$$

Remark 2.14. It is apparent that $\mathbf{S_h}$ is the s-tuple of the discrete Laplacian $-\Delta_h$, and if $\underline{u}_h = \sum c_i \underline{w}_h^i$ and $\underline{v}_h = \sum d_i \underline{w}_h^i$ then $\langle \underline{u}_h, \underline{v}_h \rangle = \mathbf{S_h c} \cdot \mathbf{d}$.

From equations (1.26) and (2.20) we have that the Newton linearization of (2.18) leads us to the linear problem: find the element $\underline{p}_h \in \mathcal{V}_h$ that satisfies

$$\int_\Omega \left(\nabla \underline{p}_h \cdot \nabla \underline{v} + \partial_{\underline{u}} f(x, \underline{u}) \underline{p}_h \cdot \underline{v} \right) = - \int_\Omega \left(\nabla \underline{u} \cdot \nabla \underline{v} + f(x, \underline{u}) \underline{v} - \underline{g} \cdot \underline{v} \right), \text{ for all } \underline{v} \in \mathcal{V}_h, \tag{2.22}$$

that is we have the following linear equation, with $\underline{p}_h = \sum_j p_j \underline{w}_h^j$ and $\mathbf{p} = (p_j)_{j=1}^m$:

$$(\mathbf{I} + \mathbf{S_h}^{-1} \mathbf{D_h}(\underline{\mathbf{u}}))\mathbf{p} = \mathbf{f}, \tag{2.23}$$

where $\mathbf{f} = \mathbf{S_h}^{-1} (\gamma_1, \ldots, \gamma_m)^T$ with

$$\gamma_j = - \int_\Omega \left(\nabla \underline{u} \cdot \nabla \underline{w}_h^j + f(x, \underline{u}) \underline{w}_h^j - \underline{g} \cdot \underline{w}_h^j \right)$$

Inner CG for the discretized equation

We see that the linear subproblem of the Newton method has the form of the linear PDE system for which we proposed the superlinear PCG algorithm.

By Proposition 2.9 F is a compact perturbation of the identity. It is then well-known that the CG method applied to (1.26) has superlinear convergence [49, 27, 9]. Moreover we have a discretization independent estimate on the convergence:

Theorem 2.15 ([27, 2]). *The PCG applied to the equation* (2.23) *yields the following convergence estimate with the notation* $\mathbf{e}_k = \mathbf{p}_k - \mathbf{p}$,*:*

$$\frac{\|\mathbf{e}_k\|_{\mathbf{S_h}+\mathbf{D_h}(\underline{\mathbf{u}})}}{\|\mathbf{e}_o\|_{\mathbf{S_h}+\mathbf{D_h}(\underline{\mathbf{u}})}} \leq \left(\frac{3 \left| \left| \left| \mathbf{S_h}^{-1} \mathbf{D_h}(\underline{\mathbf{u}}) \right| \right| \right|^2}{2k} \right)^{k/2},$$

if $k \in \mathbb{N}$ *is even and* $k \geq \frac{2}{3} \left| \left| \left| \mathbf{S_h}^{-1} \mathbf{D_h}(\underline{\mathbf{u}}) \right| \right| \right|^2$. *This estimate is independent of the subspace* $\mathcal{V}_h$ *used in Galerkin discretization.*

We can combine condition (1.4) with Theorem 2.15, thus in the nth outer iteration we need to take k_n inner iterations in order to achieve the required estimate (1.4). This means that for $n \geq 0$ k_n shall satisfy

$$\frac{\|F_h'(u_n) p_n^{k_n} + (F_h(u_n) - \underline{b}_h)\|}{\|F_h(u_n) - \underline{b}_h\|} \leq \delta_n,$$

that is with $p_n^0 = 0$ we have the estimate on k_n

$$\left(\frac{3\,|||K|||^2}{2k_n}\right)^{k_n/2} \leq \delta_n.$$

2.2.4 The algorithm

The DIN algorithm applied to the problem (2.19) is then

Initialize: calculate the matrix Δ_h (since S_h is the s-tuple of it) and set the initial guess $\mathbf{c_0} = 0$,

calculate $\mathbf{b}$ by some fast Poisson solver as a preconditioner,

calculate the residual $\mathbf{r_0} = \mathbf{c} + N_h(\mathbf{c_0}) - \mathbf{b}$, and its norm $\|\mathbf{r_0}\|_{\mathbf{S_h}}$,

while $\|r_n\| > \varepsilon$ **do**

calculate the mass matrix $\mathbf{D}(\underline{\mathbf{u}}_\mathbf{h})$ as in (2.21), and set initial value $\mathbf{p_0^0} = 0$,

calculate the residual $\mathbf{e_n^0} = \mathbf{p_n^0} + \mathbf{D}(\underline{\mathbf{u}}_\mathbf{h})\mathbf{p_n^0} - \mathbf{f}$,

calculate $\|\mathbf{e_n^0}\|_{\mathbf{S_h}}$,

define $\mathbf{q_n^0} = \mathbf{e_n^0}$,

while $\|\mathbf{e_n^k}\| > \delta_n$ **do**

calculate the constant α^k and then modify $\mathbf{p_n^k}$ and $\mathbf{e_n^k}$ as

$$\alpha^k = \frac{\mathbf{S_h e_n^k} \cdot \mathbf{q^k}}{(\mathbf{S_h} + \mathbf{D}(\underline{\mathbf{u}}_\mathbf{h}))\,\mathbf{e_n^k} \cdot \mathbf{q_n^k}}, \text{ and}$$

$\mathbf{p_n^{k+1}} = \mathbf{p_n^k} - \alpha^k \mathbf{q_n^k}$, $\mathbf{e_n^{k+1}} = \mathbf{e_n^k} - \alpha^k(\mathbf{q_n^k} + \mathbf{S_h}^{-1}\mathbf{D}(\underline{\mathbf{u}}_\mathbf{h})\mathbf{q_n^k})$ respectively,

calculate the constant β^k and then modify $\mathbf{q_n^k}$ as

$$\beta^k = \frac{(\mathbf{S_h} + \mathbf{D}(\underline{\mathbf{u}}_\mathbf{h}))\,\mathbf{e_n^{k+1}} \cdot \mathbf{q_n^k}}{(\mathbf{S_h} + \mathbf{D}(\underline{\mathbf{u}}_\mathbf{h}))\,\mathbf{e_n^k} \cdot \mathbf{q_n^k}}, \quad \mathbf{q_n^{k+1}} = \mathbf{e_n^{k+1}} - \beta^k \mathbf{q_n^k},$$

calculate the residual $\mathbf{e_n^{k+1}} = \mathbf{p_n^{k+1}} + \mathbf{D}(\underline{\mathbf{u}}_\mathbf{h})\mathbf{p_n^{k+1}} - \mathbf{f}$

calculate $\|\mathbf{e_n^{k+1}}\|_{\mathbf{S_h}}$,

end while

calculate the damping parameter τ_n and let

$$\mathbf{c_{n+1}} = \mathbf{c_n} + \tau_n \mathbf{p_n},$$

calculate the residual $\mathbf{r_n} = \mathbf{c_n} + N_h(\mathbf{c}) - \mathbf{b}$

calculate $\|\mathbf{r_n}\|_{\mathbf{S_h}}$.

end while

2.2.5 Numerical experiments

We made experiments on some test-problems below:

- the domain was $\Omega = [0,1] \times [0,1]$,
- we used Courant elements for the FEM discretization using uniform mesh with width $h = 1/N$ where N is the number of subintervals on the interval $[0,1] \times \{0\}$,
- the coordinates of the exact solutions were chosen among the functions of form $u(x,y) = C \cdot x(1-x)y(1-y)$ and $u(x,y) = C \cdot \sin \pi x \sin \pi y$,
- we had the function f as the derivative of the functional $\varphi(\xi) = \|\xi\|^4$,
- the stopping criterion was $\|F_h(\underline{u}_n) - b_h\| \leq 10^{-5}$,
- we used adaptive damping parameters τ_n,
- the code was written in Matlab.

Proposition 2.16. *The above test-problems satisfy assumptions [P1],[P2],[P3],[P4'],[P5'].*

The cases [P1]-[P3] are obvious. By some elementary calculations we have that [P4'] is satisfied with $c_1 = 0, c_2 = 12, p = 4$ and [P5'] is satisfied with $C(r) = 24r$.

Denoting $r_n = F_h(\underline{u}_n) - b_h$, n_{inn} equals the number of inner iterations, we had the following results:

Remark 2.17. The relaxing parameters τ_n defined in (1.4) produced linear convergence before the superlinear phase, but then the convergence quotient were so close to 1, that it would have needed too much computer time to reach the superlinear phase. Therefore we used some adaptive relaxing parameters.

We observe the mesh independence for both the outer and inner iterations.

Tab. 2.1: Results for $s = 2$ and $s = 6$

	$N = 25$		$N = 55$		$N = 85$	
n	$\|r_n\|$	n_{inn}	$\|r_n\|$	n_{inn}	$\|r_n\|$	n_{inn}
1	2.4747	1	2.4804	1	2.4812	1
2	1.8506	1	1.8547	1	1.8553	1
3	1.1298	1	1.1319	1	1.1323	1
4	0.4614	1	0.46195	1	0.46203	1
5	$6.2785 \cdot 10^{-2}$	2	$6.2886 \cdot 10^{-2}$	2	$6.2902 \cdot 10^{-2}$	2
6	$2.85 \cdot 10^{-4}$	3	$2.9349 \cdot 10^{-4}$	3	$2.9479 \cdot 10^{-4}$	3

	$N = 25$		$N = 35$		$N = 45$	
n	$\|r_n\|$	n_{inn}	$\|r_n\|$	n_{inn}	$\|r_n\|$	n_{inn}
1	22.294	1	22.327	1	22.341	1
2	12.422	1	12.400	1	12.390	1
3	6.9112	1	6.9049	1	6.9023	1
4	2.7724	1	2.7730	1	2.7732	1
5	1.1069	2	1.1173	2	1.1217	2
6	$1.3284 \cdot 10{-}1$	3	$1.3589 \cdot 10{-}1$	3	$1.3723 \cdot 10{-}1$	3
7	$2.4111 \cdot 10{-}3$	5	$2.4928 \cdot 10{-}3$	5	$2.5437 \cdot 10{-}3$	5
8	$8.9220 \cdot 10^{-6}$	8	$5.4236 \cdot 10^{-6}$	8	$3.6845 \cdot 10^{-6}$	8

Tab. 2.2: Results for $s = 8$

N	total number of inner PCG iterations	computer time (in *sec*)	the final error $\|\underline{u} - \underline{u}_h\|$
15	39	$2.6034 \cdot 10^{2}$	$1.4628 \cdot 10^{-1}$
25	28	$7.6434 \cdot 10^{2}$	$4.4973 \cdot 10^{-2}$
35	35	$2.3816 \cdot 10^{3}$	$2.2530 \cdot 10^{-2}$
45	23	$4.7176 \cdot 10^{3}$	$1.3578 \cdot 10^{-2}$

3. CONVERGENCE ESTIMATES USING THE MIN-MAX PRINCIPLE

The Hilbert-Schmidt estimate for the PCG algorithm used previously is an neat way of deriving superlinear estimates for the solution of elliptic systems using elegant ideas of functional analysis. The problem is that it binds us to strict conditions on both the dimension of the space where Ω lies and the parameters of the equation. As we have seen the inverse of the Laplacian is Hilbert-Schmidt only for dimensions $d \leq 3$. This is not a crucial bound as most elliptic equations in physical applications take place in these spaces. The second most pressing issue is the bounding constraints on the coefficients in the equations. One way of widing the availability of superlinear convergence is to replace the Hilbert-Schmidt norm with the so called Schatten norms. We give a schetch of this idea.

Definition 3.1 (Schatten class operator (see in e.g. [37])). Let H be separable Hilbert space, and T a compact operator in $B(H)$. For $1 \leq p < \infty$, define the Schatten p-norm of T as

$$\|T\|_p := \left(\sum_{n \geq 1} \lambda_n^p (T^* T) \right)^{1/p}.$$

For fixed p the collection of operators with finite p-norm form a Banach space. For $p = 2$ the p-norm is exactly the Hilbert-Schmidt norm.

The availability of using this comes apparent if we look back how the Frobenius norm, and so the Hilbert-Schmidt norm, came up. For a positive compact perturbation of the identity $A = I + C$ denote the eigenvalues of C in nondincreasing order $\lambda_1 \geq \lambda_2 \geq \dots$. We have the following estimate for the error $e_n = x_n - x^*$ of the CG [49].

$$\left(\frac{\|e_n\|_A}{\|e_0\|_A} \right)^{1/n} \leq \frac{2}{n} \sum_{j=1}^{n} \lambda_j \leq 2 \left(\frac{\sum_{j=1}^{n} \lambda_j^2}{n} \right)^{1/2} \leq 2 \left(\frac{|||C|||}{n} \right)^{1/2}. \tag{3.1}$$

Here we used the inequality between the arithmetic and the quadratic mean. If we replace this with a general power mean with parameter p, then we arrive to the Schatten p-norms. Thus using these norms we might consider a broader class of compact operators.

This seems a good way but it has a little technical drawback. In the following we will consider nonsymmetric equations also of the form $A = I + C$. This time the CG method is applied to the normal equations,i.e. the operator $I + C + C^* + C^*C$ is in focus. It is slightly inconvenient calculating the p-norm of the operator $C^* + C^*C$. Even more in the light of the fact that we can use an another technique. Namely we could go one step back in (3.1), before the use of the power mean inequality, and estimate the eigenvalues themself. This idea is applied in this chapter.

As applications to this method we consider a class of nonlinear nonsymmetric elliptic system of equations that is discussed in the first section. And in the second section we consider a class of nonlinear elliptic interface problems. We show that the Newton iteration has superlinear convergence and the solution algorithms of the linear subproblems has also superlinear convergence. For the linear subproblems we use preconditioned conjugate gradient method for normal equations (PCGN) and preconditionen conjugate gradient method (PCG) for the case of nonsymmetric and symmetric classes respectively.

3.1 Nonlinear nonsymmetric elliptic system of equations

In this section numerical solution of nonlinear elliptic transport systems is considered. An outer-inner (damped inexact Newton plus PCG type) iteration is proposed for the finite element discretization of the problem, and mesh independent superlinear convergence is proved for both the outer and inner iterations. Numerical experiments are enclosed.

Nonlinear elliptic transport systems arise in various problems in applied mathematics, most often leading to large-scale problems owing to the huge number of equations, see e.g. [40, 46, 50]. For large-scale elliptic problems, iterative processes are the most widespread solution methods, which often rely on Hilbert space theory when mesh independence is desired. (See e.g. [17, 30, 36] and work of Karátson [11, 18, 28].)

We consider elliptic transport systems with coupling in the nonlinear reaction terms, for which polynomial growth is allowed, and suitable coercivity is prescribed which can be naturally satisfied when the problem arises from time discretization of parabolic problems. We propose an outer-inner (damped inexact Newton plus PCGN) iteration for the finite element discretization of the problem, and prove mesh independent superlinear convergence for both the outer and inner iterations. Numerical experiments strengthen our theoretical results.

3.1.1 The problem

We consider nonlinear elliptic transport systems of the form

$$\left.\begin{aligned} -\operatorname{div}(K_i \nabla u_i) + \mathbf{b_i} \cdot \nabla \mathbf{u_i} + \mathbf{f_i}(\mathbf{x}, \mathbf{u_1}, \ldots, \mathbf{u_l}) &= \mathbf{g_i} \\ u_{i\,|\partial\Omega} &= 0 \end{aligned}\right\} \qquad (i = 1, \ldots, l) \tag{3.2}$$

on a bounded domain $\Omega \subset \mathbf{R}^d$ ($d = 2$ or 3) under the following assumptions:

Assumptions 2.1.

(i) (Smoothness:) $K_i \in L^\infty(\Omega)$, $\mathbf{b_i} \in \mathbf{C^1}(\overline{\Omega})^\mathbf{d}$ and $g_i \in L^2(\Omega)$ $(i = 1, \ldots, l)$, further, the function $f = (f_1, \ldots, f_l) : \Omega \times \mathbf{R}^l \to \mathbf{R}^l$ is measurable and bounded w.r. to the variable $x \in \Omega$ and C^1 in the variable $\xi \in \mathbf{R}^l$.

(ii) (Coercivity:) there is $m > 0$ such that $K_i \geq m$ holds for all $i = 1, \ldots, l$, further, using the notation $f'_\xi(x, \xi) := \frac{\partial f(x,\xi)}{\partial \xi}$,

$$f'_\xi(x, \xi)\, \eta \cdot \eta - \frac{1}{2} \left(\max_i \operatorname{div} \mathbf{b_i}(\mathbf{x})\right) |\eta|^2 \geq 0 \tag{3.3}$$

for any $(x, \xi) \in \Omega \times \mathbf{R}^l$ and $\eta \in \mathbf{R}^l$.

(iii) (Local Lipschitz continuity:) let $3 \leq p$ (if $d = 2$) or $3 \leq p < 6$ (if $d = 3$), then there exist constants $c_1, c_2 \geq 0$ such that for any (x, ξ_1) and $(x, \xi_2) \in \Omega \times \mathbf{R}^l$,

$$\left\| f'_\xi(x, \xi_1) - f'_\xi(x, \xi_2) \right\| \leq \left(c_1 + c_2 \left(\max |\xi_1|, |\xi_2|\right)^{p-3}\right) |\xi_1 - \xi_2|.$$

(iv) (Bounded growth:) let p be defined as in (iii), then there exist constants $c_3, c_4 > 0$ such that for any (x, ξ_1) and $(x, \xi_2) \in \Omega \times \mathbf{R}^l$,

$$|f'_\xi(x, \xi)| \leq c_3 + c_4 |\xi|^{p-2}.$$

We note that assumption 2.1, (iii) implies the estimates

$$\left\| f'_\xi(x, \xi) \right\| \leq c_3 + c_4 |\xi|^{p-2}, \qquad |f(x, \xi)| \leq c_5 + c_6 |\xi|^{p-1} \tag{3.4}$$

for any $(x, \xi) \in \Omega \times \mathbf{R}^l$.

Systems of the form (3.2) arise e.g. from the time discretization of nonlinear reaction-convection-diffusion (transport) systems

$$\left.\begin{aligned} \frac{\partial c_i}{\partial t} - \operatorname{div}(K_i \nabla c_i) + \mathbf{b_i} \cdot \nabla \mathbf{c_i} + \mathbf{R_i}(\mathbf{x}, \mathbf{c_1}, \ldots, \mathbf{c_l}) = 0 \\ c_{i\,|\partial\Omega} = 0 \end{aligned}\right\} \qquad (i = 1, \ldots, l). \tag{3.5}$$

In many real-life problems, e. g. where c_i are concentrations of chemical species, such systems may consist of a huge number of equations [50]. Using a time discretization with sufficiently small steplength τ, the obtained nonlinear elliptic systems satisfy the coercivity assumptions above.

We note that the analysis in this work remains the same when the scalar diffusion coefficients K_i in (3.2) are replaced by uniformly positive matrix coefficients. In this case the auxiliary problems (3.21) in the inner iteration have less favourable properties than for scalar coefficients, and are possibly solved by an additional inner preconditioned iteration, which we do not consider here. However, the case of scalar coefficients (or even a constant K_i) covers most of practical problems like (3.5).

For brevity, we write (3.2) as

$$\left.\begin{aligned} -\operatorname{div}(\mathbf{K} \nabla \mathbf{u}) + \mathbf{b} \cdot \nabla \mathbf{u} + f(x, \mathbf{u}) = \mathbf{g} \\ \mathbf{u}_{|\partial\Omega} = 0 \end{aligned}\right\} \tag{3.6}$$

using obvious notations.

3.1.2 Weak formulation and properties

The required theoretical background for our problem is formulated with standard Sobolev space technique. As seen before we consider the product Sobolev space as before $H_0^1(\Omega)^l := H_0^1(\Omega) \times \cdots \times H_0^1(\Omega)$ as a real Hilbert space endowed with the inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle_{H_0^1} := \sum_{i=1}^{l} \int_\Omega \nabla u_i \cdot \nabla v_i. \tag{3.7}$$

For any $\mathbf{u} \in H_0^1(\Omega)^l$ let

$$\langle F(\mathbf{u}), \mathbf{v} \rangle_{H_0^1} = \int_\Omega \sum_{i=1}^{l} \Big(K_i \, \nabla u_i \cdot \nabla v_i + (\mathbf{b_i} \cdot \nabla \mathbf{u_i})\, \mathbf{v_i} + \mathbf{f_i}(\mathbf{x}, \mathbf{u})\, \mathbf{v_i} \Big)$$

$$\equiv \int_\Omega \Big(\mathbf{K} \, \nabla \mathbf{u} \cdot \nabla \mathbf{v} + (\mathbf{b} \cdot \nabla \mathbf{u}) \cdot \mathbf{v} + f(x, \mathbf{u}) \cdot \mathbf{v} \Big) \qquad (\mathbf{v} \in H_0^1(\Omega)^l). \tag{3.8}$$

This relation defines an operator $F : H_0^1(\Omega)^l \to H_0^1(\Omega)^l$ via the Riesz representation theorem, since for any fixed $\mathbf{u} \in H_0^1(\Omega)^l$ the r.h.s. integral defines a bounded linear functional on $H_0^1(\Omega)^l$. The latter is seen in a standard way [51], using the growth condition on f in (3.4). Here we rely on the Sobolev embedding theorems Theorem 1.19, for convenience we recite them here. If $p^* := +\infty$ (if $d = 2$) or $p^* := 6$ (if $d = 3$), then for all $p \leq p^*$ we have the embedding and corresponding estimate

$$H_0^1(\Omega) \subset L^p(\Omega), \qquad \|v\|_{L^p(\Omega)} \leq C_p \cdot \|v\|_{H_0^1(\Omega)} \quad (v \in H_0^1(\Omega)). \tag{3.9}$$

with some constant $C_p > 0$.

Proposition 3.2. *The operator $F : H_0^1(\Omega)^l \to H_0^1(\Omega)^l$ is Gateaux differentiable and satisfies*

$$\langle F'(\mathbf{u})\mathbf{h}, \mathbf{h}\rangle_{H_0^1} \geq m\|\mathbf{h}\|_{H_0^1}^2 \qquad (\mathbf{u}, \mathbf{h} \in H_0^1(\Omega)^l), \tag{3.10}$$

further, F' is locally Lipschitz continuous, namely,

$$\|F'(\mathbf{u}) - F'(\mathbf{v})\| \leq L(r)\,\|\mathbf{u} - \mathbf{v}\|_{H_0^1}$$

for all $\mathbf{u}, \mathbf{v} \in H_0^1(\Omega)^l$ with $\|\mathbf{u}\|_{H_0^1} \leq r$, $\|\mathbf{v}\|_{H_0^1} \leq r$, where

$$L(r) := c_1 C_3^3 + c_2 C_p^p\, r^{p-3} \qquad (r > 0). \tag{3.11}$$

Proof.
for (3.8) it needs to be proved only for the nonlinear part: then it follows e.g. from [18]. Using the divergence theorem and assumption 2.1, (ii), we obtain

$$\langle F'(\mathbf{u})\mathbf{h}, \mathbf{h}\rangle_{H_0^1} = \int_\Omega \Big(\mathbf{K}\,|\nabla \mathbf{h}|^2 + f'_\xi(x, \mathbf{u})\,\mathbf{h}\cdot\mathbf{h} - \frac{1}{2}\sum_{i=1}^{l} (\operatorname{div} \mathbf{b}_i)\,\mathbf{h}_i^2\Big) \geq \mathbf{m}\int_\Omega |\nabla \mathbf{h}|^2\,.$$

(2) Assumption 2.1, (iii) implies for any (x, ξ_1) and $(x, \xi_2) \in \Omega \times \mathbf{R}^l$ and $\eta, \zeta \in \mathbf{R}^l$,

$$\Big|\Big(f'_\xi(x, \xi_1) - f'_\xi(x, \xi_2)\Big)\,\eta\cdot\zeta\Big| \leq \Big(c_1 + c_2\,(\max\,|\xi_1|, |\xi_2|)^{p-3}\Big)|\xi_1 - \xi_2|\,|\eta|\,|\zeta|\,,$$

hence for all $\mathbf{u}, \mathbf{v}, \mathbf{h}, \mathbf{z} \in H_0^1(\Omega)^l$

$$\Big|\langle (F'(\mathbf{u}) - F'(\mathbf{v}))\mathbf{h}, \mathbf{z}\rangle_{H_0^1}\Big| = \Big|\int_\Omega \big(f'_\xi(x, \mathbf{u}) - f'_\xi(x, \mathbf{v})\big)\,\mathbf{h}\cdot\mathbf{z}\Big|$$

$$\leq \int_\Omega \Big(c_1 + c_2\,(\max\,|\mathbf{u}|, |\mathbf{v}|)^{p-3}\Big)|\mathbf{u} - \mathbf{v}|\,|\mathbf{h}|\,|\mathbf{z}|$$

$$\leq c_1 \|\mathbf{u}-\mathbf{v}\|_{L^3} \|\mathbf{h}\|_{L^3} \|\mathbf{z}\|_{L^3} + c_2 \left(\max \|\mathbf{u}\|_{L^p}, \|\mathbf{v}\|_{L^p}\right)^{p-3} \|\mathbf{u}-\mathbf{v}\|_{L^p} \|\mathbf{h}\|_{L^p} \|\mathbf{z}\|_{L^p}$$

where for any $\mathbf{u} \in H_0^1(\Omega)^l$, $\|\mathbf{u}\|_{L^p} \equiv \|\mathbf{u}\|_{L^p(\Omega)^l} := (\int_\Omega |\mathbf{u}|^p)^{1/p}$, and in the last estimate Hölder's inequality has been used for the cases $\frac{1}{3}+\frac{1}{3}+\frac{1}{3}=1$ and $\frac{p-3}{p}+\frac{1}{p}+\frac{1}{p}+\frac{1}{p}=1$. Then (3.9) yields

$$\left|\langle (F'(\mathbf{u})-F'(\mathbf{v}))\mathbf{h}, \mathbf{z}\rangle_{H_0^1}\right|$$

$$\leq c_1 C_3^3 \|\mathbf{u}-\mathbf{v}\|_{H_0^1} \|\mathbf{h}\|_{H_0^1} \|\mathbf{z}\|_{H_0^1} + c_2 C_p^p \left(\max \|\mathbf{u}\|_{H_0^1}, \|\mathbf{v}\|_{H_0^1}\right)^{p-3} \|\mathbf{u}-\mathbf{v}\|_{H_0^1} \|\mathbf{h}\|_{H_0^1} \|\mathbf{z}\|_{H_0^1},$$

hence

$$\|F'(\mathbf{u})-F'(\mathbf{v})\| = \sup_{\substack{\mathbf{h},\mathbf{z}\in H_0^1(\Omega)^l \\ \|\mathbf{h}\|_{H_0^1}=\|\mathbf{z}\|_{H_0^1}=1}} \left|\langle (F'(\mathbf{u})-F'(\mathbf{v}))\mathbf{h}, \mathbf{z}\rangle_{H_0^1}\right|$$

$$\leq \left(c_1 C_3^3 + c_2 C_p^p \left(\max \|\mathbf{u}\|_{H_0^1}, \|\mathbf{v}\|_{H_0^1}\right)^{p-3}\right) \|\mathbf{u}-\mathbf{v}\|_{H_0^1}.$$

□

Proposition 3.3. *System (3.2) has a unique weak solution, i.e.,* $\mathbf{u} \in H_0^1(\Omega)^l$ *satisfying*

$$\langle F(\mathbf{u}), \mathbf{v}\rangle_{H_0^1} = \int_\Omega \mathbf{g}\cdot\mathbf{v} \qquad (\mathbf{v} \in H_0^1(\Omega)^l).$$

Proof. The coercivity (3.10) implies that for all $\mathbf{u} \in H_0^1(\Omega)^l$ the operator $F'(\mathbf{u})$ is regular, i.e. maps onto $H_0^1(\Omega)^l$, further,

$$\|F'(\mathbf{u})\mathbf{h}\|_{H_0^1} \geq m\|\mathbf{h}\|_{H_0^1} \qquad (\mathbf{h} \in H_0^1(\Omega)^l). \tag{3.12}$$

Then again a Hadamard type theorem [39] ensures a unique solution for the rquation $F(\mathbf{u}) = \mathbf{f}$ defined as

$$\langle F(\mathbf{u}), \mathbf{v}\rangle_{H_0^1} = \langle \mathbf{f}, \mathbf{v}\rangle_{H_0^1} \qquad (\mathbf{v} \in H_0^1(\Omega)^l)$$

where the vector $\mathbf{f} \in H_0^1(\Omega)^l$ satisfies $\langle \mathbf{f}, \mathbf{v}\rangle_{H_0^1} = \int_\Omega \mathbf{g}\cdot\mathbf{v}$ $(\mathbf{v} \in H_0^1(\Omega)^l)$, and the existence of $\mathbf{f}$ follows from the Riesz representation theorem. □

3.1.3 FEM discretization and Newton iteration

Let us consider the FEM discretization of system (3.6) in some FEM subspace

$$V_h = \operatorname{span}\{\varphi_1, \ldots, \varphi_N\} \subset H_0^1,$$

where φ_i are linearly independent. We seek the FEM solution $\mathbf{u}_h \in V_h$:

$$\langle F(\mathbf{u}_h), \mathbf{v}_h \rangle_{H_0^1} = \int_\Omega \mathbf{g} \cdot \mathbf{v}_h \qquad (\mathbf{v} \in V_h).$$

Defining the operator $F_h : V_h \to V_h$ and the function $\mathbf{g}_h \in V_h$ by the identities $\langle F_h(\mathbf{u}_h), \mathbf{v}_h \rangle_{H_0^1} = \langle F(\mathbf{u}_h), \mathbf{v}_h \rangle_{H_0^1}$ $(\forall \mathbf{v} \in V_h)$ and $\langle \mathbf{g}_h, \mathbf{v}_h \rangle_{H_0^1} = \int_\Omega \mathbf{g} \cdot \mathbf{v}_h$ $(\forall \mathbf{v} \in V_h)$, respectively, we can write our problem as

$$F_h(\mathbf{u}_h) = \mathbf{g}_h \tag{3.13}$$

in V_h, which requires the solution of an $N \times N$ nonlinear algebraic system for the coefficient vector of $\mathbf{u}_h$.

We apply the damped inexact Newton method (DIN) for the iterative solution of problem (3.13) which posesses the following convergence property:

Theorem 3.4. *Let Assumptions 2.1 hold. Let $\mathbf{u}_0 \in V_h$ be arbitrary, and let us define $R_0 := 2m^{-1}\|F_h(\mathbf{u}_0) - \mathbf{g}_h\|_{H_0^1} + \|\mathbf{u}_0\|_{H_0^1}$ and $L := L(R_0)$ with the function $L(r)$ defined in (3.11). The DIN iteration Theorem 1.25 defines a sequence $(\mathbf{u}_n) \subset V_h$*

Proof. We have seen in the proof of Proposition 3.3 that for all $\mathbf{u} \in H_0^1(\Omega)^l$ the operator $F'(\mathbf{u})$ is regular and (3.12) holds. Further, by Proposition 3.2, F' is locally Lipschitz continuous. These properties are inherited with the same constants by the operator F_h in V_h by definition, and they imply the given convergence estimates of the DIN method (see e.g. [18], Theorem 5.12 and Remark 5.17). In particular, as pointed out in the cited remark, u_n satisfies the a priori estimate

$$\|\mathbf{u}_n\|_{H_0^1} \le R_0 \tag{3.14}$$

with R_0 given in the theorem, hence the formulation involves the global Lipschitz constant $L := L(R_0)$. □

Remark 3.5. (Mesh independence.) Let $r_n := \|F_h(\mathbf{u}_n) - \mathbf{g}_h\|_{H_0^1}$. As shown by [18, Theorem 5.12], the linear convergence factor for the first n_0 steps depends on L, m, r_0 and δ_0, whereas d_1 and q in the superlinear estimate in Theorem 1.25 depends on L, m, r_{n_0} and the prescribed sequence δ_n. If, for a sequence of FEM subspaces V_h such that $h \to 0$, we define $\mathbf{u}_0 \in V_h$ as the projection of a fixed function in $H_0^1(\Omega)^l$, e.g. $\mathbf{u}_0 := 0$, then r_0 is bounded in h and the other constants L, m, δ_0 and δ_n are given independently of h. Further, r_{n_0} can be prescribed by the choice of the step when we start the undamped part of the iteration. Hence the convergence rate of the DIN iteration is bounded mesh independently.

Remark 3.6. The value of τ_n in Theorem 3.4 uses (3.11) and shows that as $\|F_h(\mathbf{u}_n) - \mathbf{g}_h\|_1 \to 0$, the steplength τ_n reaches its optimal value 1 which holds for an undamped Newton step. In practice the above value serves as a bound on the steplength, and in actual computations most often other techniques like adaptive updating are used to compute efficient steplengths. See e.g. [7] for further discussion.

3.1.4 *Solution of the linearized problems: inner CG type iterations*

Let $\mathbf{u}_n$ be constructed in the DIN iteration, and let us consider the linearized problem

$$\mathbf{F}'_h(\mathbf{u}_n)\mathbf{p}_h = \mathbf{r}_h \tag{3.15}$$

(where $\mathbf{r}_h := \mathbf{g}_h - \mathbf{F}_h(\mathbf{u}_n)$), which is equivalent to the FEM solution in V_h of the linear elliptic problem

$$\left.\begin{aligned} -\operatorname{div}(K_i \nabla p_i) + \mathbf{b_i} \cdot \nabla \mathbf{p_i} + \sum_{\mathbf{j}=1}^{\mathbf{l}} \partial_{\mathbf{j}} \mathbf{f_i}(\mathbf{x}, \mathbf{u_n}) \mathbf{p_j} = \mathbf{r_i} \\ p_{i\,|\partial\Omega} = 0 \end{aligned}\right\} \qquad (i = 1, \ldots, l) \tag{3.16}$$

where $r_i = g_i + \operatorname{div}(K_i \nabla u_{n,i}) - \mathbf{b_i} \cdot \nabla \mathbf{u_{n,i}} - \mathbf{f_i}(\mathbf{x}, \mathbf{u_n})$. Denoting by $\mathbf{c}$ and $\mathbf{d}$ the coefficient vectors of $\mathbf{p}_h$ and $\mathbf{r}_h$, respectively, and by $\mathbf{L}_h^{(n)}$ the stiffness matrix corresponding to the linear problem (3.16), we need to solve the linear algebraic system

$$\mathbf{L}_h^{(n)} \mathbf{c} = \mathbf{d}. \tag{3.17}$$

We propose a preconditioned conjugate gradient method to solve (3.17). We define our preconditioners based on the following equivalent operator: letting

$$S_i u_i := -\operatorname{div}(K_i \nabla u_i) + h_i u_i \qquad (i = 1, \ldots, l) \tag{3.18}$$

(for $u_{i\,|\partial\Omega} = 0$), where $h_i \in L^\infty(\Omega)$ and $h_i \geq 0$, we define the independent l-tuple of elliptic operators

$$S\mathbf{u} = (S_1 u_1, \ \ldots, \ S_l u_l). \tag{3.19}$$

We consider the preconditioned form of the algebraic system (3.17):

$$\mathbf{S}_h^{-1} \mathbf{L}_h^{(n)} \mathbf{c} = \mathbf{f} \tag{3.20}$$

(with $\mathbf{f} = \mathbf{S}_h^{-1}\mathbf{c}$), where $\mathbf{S}_h$ denotes the stiffness matrix of S in the same FEM subspace V_h. This preconditioning leads to the FEM solutions in V_h of independent symmetric

auxiliary linear elliptic problems of the form

$$-\mathrm{div}\,(K_i \nabla z_i) + h_i z_i = f_i \qquad (i = 1, \ldots, l). \tag{3.21}$$

For such problems various fast solvers are available that turn $\mathbf{S}_h$ into an efficient preconditioner, in particular when K_i and h_i are also constant. Optimal order linear solvers like multigrid and multilevel are proposed above all [8, 20], further, on special domains, FFT or cyclic reduction can be considered [42, 45]).

Our goal is to apply a suitable CG type iteration to (3.20).

Conjugate gradient algorithms for nonsymmetric linear problems

In this section we summarize the required results on the conjugate gradient method based on [10]. Let us consider a nonsymmetric linear algebraic system

$$Au = b \tag{3.22}$$

with given $A \in \mathbf{R}^{N\times N}$, $b \in \mathbf{R}^N$. Let $\langle .,. \rangle$ be a given inner product on $\mathbf{R}^N$ and, denoting by A^* the adjoint of A w.r.t. this inner product, assume that

$$A + A^* > 0. \tag{3.23}$$

There exist several CG algorithms for such nonsymmetric systems (see e.g. [6, 16]). One of the most widespread ways, often called CGN method (in literature it is also often called CGNE), is to consider the normal (or symmetrized) equation and apply a symmetric CG method. The algorithm itself was described in Definition 1.23.

Let us consider the decomposition $A = I + C$. Using the notation $\nu := \min_{x\in\mathbf{R}^N} \frac{\|Ax\|^2}{\|x\|^2}$, the error vector $r_k := Au_k - b$ satisfies

$$\left(\frac{\|r_k\|}{\|r_0\|}\right)^{1/k} \le \frac{2}{k\nu} \sum_{i=1}^{k} \Big(\big|\lambda_i(C^* + C)\big| + \lambda_i(C^*C) \Big) \qquad (k = 1, 2, ..., N). \tag{3.24}$$

Remark 3.7. The above results hold in Hilbert space settings also, when C is a compact operator. We may replace ν with $\|A^{-1}\|$.

The above result has a mesh independent bound when suitably applied to elliptic systems. Let us consider the Dirichlet problem

$$\left.\begin{aligned} L_i u \equiv\ & -\mathrm{div}\,(K_i \nabla u_i) + \mathbf{b_i} \cdot \nabla \mathbf{u_i} + \sum_{\mathbf{j}=1}^{\mathbf{l}} \mathbf{V_{ij}} \mathbf{u_j} = \mathbf{g_i} \\ & u_{i\,|\partial\Omega} = 0 \end{aligned}\right\} \qquad (i = 1, \ldots, l) \tag{3.25}$$

on a bounded domain $\Omega \subset \mathbf{R}^d$, where K_i is as in Assumptions 2.1, $\mathbf{b_i} \in \mathbf{C^1}(\overline{\Omega})^{\mathbf{d}}$, $g_i \in L^2(\Omega)$, $V_{ij} \in L^\infty(\Omega)$, and we assume that $\mathbf{b_i}$ and the matrix $V = \{V_{ij}\}_{i,j=1}^{l}$ satisfy the coercivity property

$$\lambda_{min}(V + V^T) - \max_i \operatorname{div} \mathbf{b_i} \geq 0 \tag{3.26}$$

pointwise on Ω, where λ_{min} denotes the smallest eigenvalue. (Then system (3.25) has a unique weak solution $u \in H_0^1(\Omega)^l$.) Let us choose a FEM subspace $V_h = \operatorname{span}\{\varphi_1, \ldots, \varphi_N\} \subset H_0^1(\Omega)^l$ and look for the solution of the corresponding algebraic system $\mathbf{L}_h\,\mathbf{c} = \mathbf{b}$. We define the preconditioning operator (3.19) and the corresponding inner product on $H_0^1(\Omega)^l$

$$\langle \mathbf{u}, \mathbf{v} \rangle_S := \int_\Omega \sum_{i=1}^{l} \Big(K_i\ \nabla u_i \cdot \nabla v_i + h_i u_i v_i \Big)$$

which is equivalent to (3.7). We propose the stiffness matrix $\mathbf{S}_h$ of S in V_h as preconditioner for system $\mathbf{L}_h\,\mathbf{c} = \mathbf{b}$, and solve the preconditioned system $\mathbf{S}_h^{-1}\mathbf{L}_h\,\mathbf{c} = \mathbf{S}_h^{-1}\mathbf{b}$ using the CG algorithm with the $\mathbf{S}_h$-inner product and with the cast $A = \mathbf{S}_h^{-1}\mathbf{L}_h$ and $A^* = \mathbf{S}_h^{-1}\mathbf{L}_h^T$. Then the following mesh independent superlinear convergence result holds, given in terms of the compact operator Q_S defined via

$$\langle Q_S \mathbf{u}, \mathbf{v} \rangle_S = \sum_{i=1}^{l} \int_\Omega \Big((\mathbf{b_i} \cdot \nabla \mathbf{u_i})\, \mathbf{v_i} + (\sum_{\mathbf{j}=1}^{\mathbf{l}} \mathbf{V_{ij}} \mathbf{u_j} - \mathbf{h_i} \mathbf{u_i})\, \mathbf{v_i} \Big) \equiv \int_\Omega \Big((\mathbf{b} \cdot \nabla \mathbf{u}) \cdot \mathbf{v} + (\mathbf{V} - \mathbf{hI})\mathbf{u} \cdot \mathbf{v} \Big) \tag{3.27}$$

($\mathbf{u}, \mathbf{v} \in H_0^1(\Omega)^l$), and denoting by $s_i(Q_S) := \lambda_i(Q_S^* Q_S)^{1/2}$ and $\lambda_i(Q_S^* + Q_S)$ $(i = 1, 2, \ldots)$ the singular values resp. ordered eigenvalues of the corresponding operators:

Theorem 3.8. [10]. *The CGN algorithm (Definition 1.23) with $\mathbf{S}_h$-inner product, applied for the $N \times N$ preconditioned system $\mathbf{S}_h^{-1}\mathbf{L}_h\,\mathbf{c} = \mathbf{S}_h^{-1}\mathbf{b}$, yields*

$$\left(\frac{\|r_k\|_{\mathbf{S}_h}}{\|r_0\|_{\mathbf{S}_h}} \right)^{1/k} \leq \varepsilon_k \qquad (k = 1, 2, \ldots, N) \tag{3.28}$$

$$\textit{where} \quad \varepsilon_k = \frac{2}{km^2} \sum_{i=1}^{k} \Big(\big|\lambda_i(Q_S^* + Q_S)\big| + s_i(Q_S)^2 \Big) \to 0 \qquad (\textit{as } k \to \infty) \tag{3.29}$$

and $(\varepsilon_k)_{k \in \mathbf{N}^+}$ is a sequence independent of n and V_h.

We note that the use of the normal equation to derive the above CGN algorithm is favourable in spite of the related increase of the condition number. Namely, the latter only influences the linear convergence bound, whereas in our situation the superlinear

convergence rate (3.29) (and, moreover, the magnitude (3.41) later) is comparable to the case when the normal equation can be avoided [11].

Uniform superlinear convergence of the inner PCGN iteration

Based on the previous subsection, we apply the CGN algorithm with $\mathbf{S}_h$-inner product to the preconditioned system (3.20). We verify that the superlinear convergence rate of this algorithm is bounded uniformly w.r.t. both the mesh and the outer Newton iterate, i.e., the sequence ε_k in (3.28) can be replaced by a sequence $\widehat{\varepsilon}_k$ which is independent of both V_h and $\mathbf{c}_n$.

We rely on Theorem 3.8. Here the operator Q_S in (3.27) now contains the Jacobian $V = f'_\xi(x, \mathbf{u}_n)$, that is, $Q_S = Q_S^{(n)}$ defined by

$$\langle Q_S^{(n)}\mathbf{v}, \mathbf{z}\rangle_S = \sum_{i=1}^{l} \int_\Omega \Big((\mathbf{b}_i \cdot \nabla \mathbf{v}_i)\, \mathbf{z}_i + (\sum_{\mathbf{j}=1}^{\mathbf{l}} \partial_{\mathbf{j}} \mathbf{f}_i(\mathbf{x}, \mathbf{u}_n)\mathbf{v}_j - \mathbf{h}_i \mathbf{v}_i)\, \mathbf{z}_i \Big)$$

$$\equiv \int_\Omega \Big((\mathbf{b} \cdot \nabla \mathbf{v}) \cdot \mathbf{z} + (f'_\xi(x, \mathbf{u}_n) - \mathbf{h}I)\mathbf{v} \cdot \mathbf{z} \Big) \qquad (\mathbf{v}, \mathbf{z} \in H_0^1(\Omega)^l).$$

Although Theorem 3.8 itself states mesh independence for the linear problem (3.25), our linearized algebraic system (3.20) depends on an outer Newton iterate $\mathbf{u}_n$ constructed in a given FEM subspace. Hence even the mesh independence part itself of the following theorem does not obviously follow from Theorem 3.8. We now give our estimate involving two minimax ratios, related to the L^2 and L^p norms, respectively.

Now we prove a simple lemma that we will use in the next theorem to prove superlinear convergence.

Lemma 3.9. *Let $C \in B(H, X)$ be a compact operator from a separable Hilbert space H to a Banach space X. Then the sequence (λ_n) defined as*

$$\lambda_n = \min_{H_{n-1} \subset H} \max_{x \perp H_{n-1}} \frac{\|Cx\|_X}{\|x\|_H}$$

converges to 0.

Proof. From now on we drop the notation of the different notation of norms on X and H, as it is unambiguous which one is used.

Let $e_1 \in H$ is a unit vector where C attains its norm. By definition $\lambda_1 = \|Ce_1\|$.

Now recursively we define the orthonormal system $(e_n) \subset H$. If $e_1, \ldots, e_n$ is already defined, define the subspace $\widehat{H}_n = \operatorname{span}\{e_1, \ldots, e_n\}$. Now define e_{n+1} as the vector satisfying

$$\|Ce_{n+1}\| = \max_{x \perp \widehat{H}_n} \frac{\|Cx\|}{\|x\|},$$

i.e. the operator C restricted to the orthocomplement of $\widehat{H}_n$ attains its norm at e_{n+1}.

By the definition of the sequence (λ_n) and the definition of (e_n) we have that

$$\lambda_n \leq \|Ce_n\|.$$

Now as the system (e_n) is orthonormal it weakly convergenes to 0, hence by the compactness of C we have that $\|Ce_n\| \to 0$ and so $\lambda_n \to 0$. □

Theorem 3.10. *The CGN algorithm (Definition 1.23) with $\mathbf{S}_h$-inner product, applied for the $N \times N$ preconditioned system (3.20), yields*

$$\left(\frac{\|r_k\|_{\mathbf{S}_h}}{\|r_0\|_{\mathbf{S}_h}}\right)^{1/k} \leq \widehat{\varepsilon}_k \qquad (k = 1, 2, ..., N) \tag{3.30}$$

with $\widehat{\varepsilon}_k :=$

$$\frac{2}{km^2} \sum_{i=1}^{k} \Big(C_1 \min_{H_{i-1} \subset H_0^1(\Omega)^l} \max_{\mathbf{v} \perp H_{i-1}} \frac{\|\mathbf{v}\|^2_{L^2(\Omega)^l}}{\|\mathbf{v}\|^2_S} + C_2 \min_{H_{i-1} \subset H_0^1(\Omega)^l} \max_{\mathbf{v} \perp H_{i-1}} \frac{\|\mathbf{v}\|^2_{L^p(\Omega)^l}}{\|\mathbf{v}\|^2_S} \Big). \tag{3.31}$$

The expression $\widehat{\varepsilon}_k$ tends to zero as $k \to \infty$ (where H_{i-1} stands for an arbitrary $(i-1)$-dimensional subspace and orthogonality is understood in S-inner product), and here the constants $C_1, C_2 > 0$ and hence the sequence $(\widehat{\varepsilon}_k)_{k \in \mathbf{N}^+}$ are independent of V_h and $\mathbf{u}_n$.

Proof. We rely on Theorem 3.8 and prove that the sequence ε_k in (3.28)-(3.29) satisfies $\varepsilon_k \leq \widehat{\varepsilon}_k$ if $Q_S = Q_S^{(n)}$ as above, further, that $\widehat{\varepsilon}_k \to 0$. The divergence theorem yields for $\mathbf{v}, \mathbf{z} \in H_0^1(\Omega)^l$

$$\int_\Omega (\mathbf{b}_i \cdot \nabla \mathbf{v}_i) \mathbf{z}_i = -\int_\Omega \mathbf{v}_i (\mathbf{b}_i \cdot \nabla \mathbf{z}_i) - \int_\Omega (\operatorname{div} \mathbf{b}_i) \mathbf{v}_i \mathbf{z}_i \,, \tag{3.32}$$

hence from (3.27) and (3.4)

$$\|Q_S^{(n)} \mathbf{v}\|_S = \sup_{\substack{\mathbf{z} \in H_0^1(\Omega)^l \\ \|\mathbf{z}\|_S = 1}} |\langle Q_S^{(n)} \mathbf{v}, \mathbf{z} \rangle_S|$$

$$= \sup_{\substack{\mathbf{z} \in H_0^1(\Omega)^l \\ \|\mathbf{z}\|_S = 1}} \left| \sum_{i=1}^{l} \int_\Omega \Big(-v_i \left(\mathbf{b}_i \cdot \nabla \mathbf{z}_i\right) + \Big(\sum_{\mathbf{j}=\mathbf{1}}^{\mathbf{l}} \partial_{\mathbf{j}} \mathbf{f}_{\mathbf{i}}(\mathbf{x}, \mathbf{u}_n) \mathbf{v}_{\mathbf{j}} - \mathbf{h}_{\mathbf{i}} \mathbf{v}_{\mathbf{i}} - (\operatorname{div} \mathbf{b}_{\mathbf{i}})\, \mathbf{v}_{\mathbf{i}} \Big)\, \mathbf{z}_{\mathbf{i}} \Big) \right|$$

$$\equiv \sup_{\substack{\mathbf{z} \in H_0^1(\Omega)^l \\ \|\mathbf{z}\|_S = 1}} \left| \int_\Omega \Big(-\mathbf{v} \cdot (\mathbf{b} \cdot \nabla \mathbf{z}) + \big(f'_\xi(x, \mathbf{u}_n) - (\mathbf{h} + \operatorname{div} \mathbf{b}) I \big) \mathbf{v} \cdot \mathbf{z} \Big) \right|$$

$$\leq \sup_{\substack{\mathbf{z}\in H_0^1(\Omega)^l \\ \|\mathbf{z}\|_S=1}} \left(\max_i \|\mathbf{b_i}\|_{\mathbf{L}^\infty(\Omega)^1} \int_\Omega |\mathbf{v}|\,|\nabla \mathbf{z}| + \left(\mathbf{c_3} + \max_{\mathbf{i}} \left\|\mathbf{h_i} + \operatorname{div} \mathbf{b_i}\right\|_{\mathbf{L}^\infty(\Omega)}\right) \int_\Omega |\mathbf{v}\mathbf{z}| \right.$$

$$\left. + c_4 \int_\Omega |\mathbf{u}_n|^{p-2} |\mathbf{v}\mathbf{z}| \right) \tag{3.33}$$

$$\leq \sup_{\substack{\mathbf{z}\in H_0^1(\Omega)^l \\ \|\mathbf{z}\|_S=1}} \left(\max_i \|\mathbf{b_i}\|_{\mathbf{L}^\infty(\Omega)^1} \|\mathbf{v}\|_{\mathbf{L}^2(\Omega)^1} \|\nabla \mathbf{z}\|_{\mathbf{L}^2(\Omega)^{1d}} + \left(\mathbf{c_3} + \max_{\mathbf{i}} \left\|\mathbf{h_i} + \operatorname{div} \mathbf{b_i}\right\|_{\mathbf{L}^\infty(\Omega)}\right) \|\mathbf{v}\|_{\mathbf{L}^2(\Omega)^1} \|\mathbf{z}\|_{\mathbf{L}^2(\Omega)^1} \right.$$

$$\left. + c_4 \|\mathbf{u}_n\|_{L^p(\Omega)^l}^{p-2} \|\mathbf{v}\|_{L^p(\Omega)^l} \|\mathbf{z}\|_{L^p(\Omega)^l} \right), \tag{3.34}$$

where in the last term Hölder's inequality has been used for the case $\frac{p-2}{p} + \frac{1}{p} + \frac{1}{p} = 1$. Here we have $\|\nabla \mathbf{z}\|_{L^2(\Omega)^{ld}} = \|\mathbf{z}\|_{H_0^1} \leq \frac{1}{\sqrt{m}} \cdot \|\mathbf{z}\|_S = \frac{1}{\sqrt{m}}$ and, also using (3.9), $\|\mathbf{z}\|_{L^p(\Omega)^l} \leq \frac{C_p}{\sqrt{m}} \cdot \|\mathbf{z}\|_S = \frac{C_p}{\sqrt{m}}$ for all $p \leq p^*$. Therefore

$$\|Q_S^{(n)} \mathbf{v}\|_S \leq \left(\frac{1}{\sqrt{m}} \max_i \|\mathbf{b_i}\|_{\mathbf{L}^\infty(\Omega)^1} \|\mathbf{v}\|_{\mathbf{L}^2(\Omega)^1} \right.$$

$$\left. + \frac{C_2}{\sqrt{m}} \left(c_3 + \max_i \left\|h_i + \operatorname{div} \mathbf{b_i}\right\|_{\mathbf{L}^\infty(\Omega)}\right) \|\mathbf{v}\|_{\mathbf{L}^2(\Omega)^1} + \mathbf{c_4} \frac{\mathbf{C_p}}{\sqrt{\mathbf{m}}} \|\mathbf{u_n}\|_{\mathbf{L^p}(\Omega)^1}^{\mathbf{p-2}} \|\mathbf{v}\|_{\mathbf{L^p}(\Omega)^1} \right),$$

moreover, from (3.9) and (3.14)

$$\|\mathbf{u}_n\|_{L^p(\Omega)^l} \leq C_p \cdot \|\mathbf{u}_n\|_{H_0^1} \leq C_p R_0\,, \tag{3.35}$$

hence

$$\|Q_S^{(n)} \mathbf{v}\|_S \leq const. \cdot \|\mathbf{v}\|_{L^2(\Omega)^l} + const. \cdot \|\mathbf{v}\|_{L^p(\Omega)^l},$$

which implies

$$\|Q_S^{(n)} \mathbf{v}\|_S^2 \leq K_1 \|\mathbf{v}\|_{L^2(\Omega)^l}^2 + K_2 \|\mathbf{v}\|_{L^p(\Omega)^l}^2 \tag{3.36}$$

and here K_1, K_2 are independent of V_h and $\mathbf{u}_n$.

Now setting $v_i = z_i$ in (3.32),

$$\int_\Omega (\mathbf{b_i} \cdot \nabla \mathbf{v_i})\, \mathbf{v_i} = -\int_\Omega \frac{\mathbf{1}}{\mathbf{2}} (\operatorname{div} \mathbf{b_i})\, \mathbf{v_i^2}$$

hence by (iv)

$$\left| \langle Q_S^{(n)} \mathbf{v}, \mathbf{v} \rangle_S \right| = \left| \sum_{i=1}^{l} \int_\Omega \left((\mathbf{b_i} \cdot \nabla \mathbf{v_i})\, \mathbf{v_i} + \Big(\sum_{\mathbf{j}=1}^{\mathbf{l}} \partial_{\mathbf{j}} \mathbf{f_i}(\mathbf{x}, \mathbf{u_n}) \mathbf{v_j} - \mathbf{h_i} \mathbf{v_i}\Big)\, \mathbf{v_i} \right) \right|$$

$$\equiv \left| \int_\Omega \Big((\mathbf{b} \cdot \nabla \mathbf{v}) \cdot \mathbf{v} + (f'_\xi(x, \mathbf{u}_n) - \mathbf{h} I) \mathbf{v} \cdot \mathbf{v} \Big) \right|$$

$$\leq \int_\Omega \max_i \left|h_i + \frac{1}{2}\operatorname{div} \mathbf{b_i}\right| |\mathbf{v}|^2 + \int_\Omega \left(\mathbf{c_3} + \mathbf{c_4}|\mathbf{u_n}|^{\mathbf{p-2}}\right) |\mathbf{v}|^2$$

$$\leq \left(c_3 + \max_i \left\|h_i + \frac{1}{2}\operatorname{div} \mathbf{b_i}\right\|_{\mathbf{L^\infty(\Omega)}}\right) \|\mathbf{v}\|^{\mathbf{2}}_{\mathbf{L^2(\Omega)^l}} + \mathbf{c_4}\|\mathbf{u_n}\|^{\mathbf{p-2}}_{\mathbf{L^p(\Omega)^l}} \|\mathbf{v}\|^{\mathbf{2}}_{\mathbf{L^p(\Omega)^l}}.$$

Using (3.35) again, we obtain

$$\left|\langle Q_S^{(n)}\mathbf{v}, \mathbf{v}\rangle_S\right| \leq K_3\|\mathbf{v}\|^2_{L^2(\Omega)^l} + K_4\|\mathbf{v}\|^2_{L^p(\Omega)^l} \tag{3.37}$$

and here K_3, K_4 are independent of h and $\mathbf{u}_n$. Now let $H_S = H_0^1(\Omega)^l$ with the S-inner product. The variational characterization of the eigenvalues yields

$$|\lambda_i\Big(\big(Q_S^{(n)}\big)^* + Q_S^{(n)}\Big)| =$$

$$\min_{H_{i-1}\subset H_S} \max_{\mathbf{v}\perp H_{i-1}} \frac{\left|\langle\Big(\big(Q_S^{(n)}\big)^* + Q_S^{(n)}\Big)\mathbf{v}, \mathbf{v}\rangle_S\right|}{\|\mathbf{v}\|_S^2} = 2 \min_{H_{i-1}\subset H_S} \max_{\mathbf{v}\perp H_{i-1}} \frac{\left|\langle Q_S^{(n)}\mathbf{v}, \mathbf{v}\rangle_S\right|}{\|\mathbf{v}\|_S^2}$$

and

$$s_i(Q_S^{(n)})^2 = \lambda_i\big(\big(Q_S^{(n)}\big)^* Q_S^{(n)}\big) =$$

$$\min_{H_{i-1}\subset H_S} \max_{\mathbf{v}\perp H_{i-1}} \frac{\langle\big(Q_S^{(n)}\big)^* Q_S^{(n)}\mathbf{v}, \mathbf{v}\rangle_S}{\|\mathbf{v}\|_S^2} \min_{H_{i-1}\subset H_S} \max_{\mathbf{v}\perp H_{i-1}} \frac{\|Q_S^{(n)}\mathbf{v}\|_S^2}{\|\mathbf{v}\|_S^2},$$

where H_{i-1} stands for an arbitrary $(i-1)$-dimensional subspace. Summing up and using (3.37) and (3.36), respectively, we obtain

$$|\lambda_i\Big(\big(Q_S^{(n)}\big)^* + Q_S^{(n)}\Big)| + s_i(Q_S^{(n)})^2 \leq$$

$$C_1 \min_{H_{i-1}\subset H_S} \max_{\mathbf{v}\perp H_{i-1}} \frac{\|\mathbf{v}\|^2_{L^2(\Omega)^l}}{\|\mathbf{v}\|_S^2} + C_2 \min_{H_{i-1}\subset H_S} \max_{\mathbf{v}\perp H_{i-1}} \frac{\|\mathbf{v}\|^2_{L^p(\Omega)^l}}{\|\mathbf{v}\|_S^2} \tag{3.38}$$

where $C_1 = 2K_3 + K_1$, $C_2 = 2K_4 + K_2$. Here both terms on the r.h.s. tend to 0 as $i \to \infty$, owing to the compactness of the embeddings $H_0^1(\Omega)^l \subset L^2(\Omega)^l$ and $H_0^1(\Omega)^l \subset L^p(\Omega)^l$ and the use of Lemma 3.9. (In particular, the first min-max term gives the reciprocal of the eigenvalues of S in $L^2(\Omega)^l$.) That is, the sequence $(\widehat{\varepsilon}_k)$ is constant times the arithmetic means of a sequence that tends to zero, hence, as is well-known, $\widehat{\varepsilon}_k$ itself tends to zero. □

Explicit asymptotics

The functions $\mathbf{u}_n \in V_h$ $(n \in \mathbf{N}^+)$ and $\mathbf{u}_h \in V_h$ are bounded since they are piecewise polynomials. If they are also uniformly bounded as $h \to 0$, which follows e.g. in the case of uniform convergence, then the term (3.33) can be estimated by

$c_4\big(\sup \|\mathbf{u}_n\|^{p-2}_{L^\infty(\Omega)^l}\big)\, \|\mathbf{v}\|_{L^2(\Omega)^l}\, \|\mathbf{z}\|_{L^2(\Omega)^l}$ instead of the Hölder estimate (3.34), i.e. this term can also be included in the L^2-norm estimates before, and (3.36) is simply replaced by

$$\|Q_S^{(n)}\mathbf{v}\|_S^2 \le K_1' \|\mathbf{v}\|^2_{L^2(\Omega)^l} \tag{3.39}$$

where the constant K_1' is independent of h and $\mathbf{u}_n$. In just the same way the L^p-norm can be eliminated from (3.36) too. Then the estimate (3.31) in Theorem 3.10 is replaced by

$$\widehat{\varepsilon}_k := \frac{C}{k}\sum_{i=1}^{k} \varrho_i\,, \qquad \text{where} \quad \varrho_i := \min_{H_{i-1}\subset H_0^1(\Omega)^l} \max_{\mathbf{v}\perp H_{i-1}} \frac{\|\mathbf{v}\|^2_{L^2(\Omega)^l}}{\|\mathbf{v}\|_S^2} \tag{3.40}$$

(where the constant $C > 0$ is independent of V_h and $\mathbf{u}_n$). Under our Dirichlet boundary conditions, as pointed out in [10], the numbers ϱ_i are the reciprocals of the eigenvalues of S for which $\varrho_i = O(i^{-2/d})$ holds [14], hence by an elementary calculation

$$\widehat{\varepsilon}_k \le O\Big(\frac{\log k}{k}\Big) \quad \text{if } d=2 \quad \text{and} \quad \widehat{\varepsilon}_k \le O\Big(\frac{1}{k^{2/3}}\Big) \quad \text{if } d=3. \tag{3.41}$$

With a bit deeper insight to the properties of the Sobolev embeddings we may obtain explicit order of convergence without further constraints. The tool to achieve such estimate is the notion of Gelfand numbers [38, 47]. Gelfand numbers, along with Kolmogorov numbers and approximation numbers were introduced when the quality of compact operators (most prominently embeddings of Sobolev spaces, Besov spaces,..) came of interest. We only cite the result that serves our need.

Definition 3.11. [38, 47] Let X, Y be two Banach spaces and let $T \in B(X,Y)$.

For $n \in \mathbb{N}$, we define the nth Gelfand number by

$$c_n(T) = \inf\{\|TJ_M^X\| : M \subset X, codim(M) < n\}.$$

Here, J_M^X stands for the natural injection of M into X.

Theorem 3.12. *[38, 47] The Gelfand numbers of the Sobolev embedding $H^1(\Omega) \hookrightarrow L^q(\Omega)$, with $\Omega \subset \mathbb{R}^d$, has the following property*

$$c_n \sim O\left(\frac{1}{n^{\frac{1}{d}-\frac{q-2}{2q}}}\right) \quad \textit{for} \quad q \le \frac{2d}{d-2}.$$

Proposition 3.13. *As done previously, we examine* (3.38). *Same as before the first minmax expression is exactly the ith eigenvalue of the inverse of S. Next we observe that the second minmax expression is exactly the ith Gelfand number of the embedding*

$H^1(\Omega)^l \hookrightarrow L^p(\Omega)^l$. *Therefore using Theorem* 3.12 *and elementary calculation shows that* $\widehat{\varepsilon}_k$ *has the order*

$$O\left(\frac{\log k}{k} + k^{-1/p}\right) = O\left(k^{-1/p}\right) \quad \text{and} \quad O\left(k^{-\frac{2}{3}} + k^{\frac{6-p}{6p}}\right) = O\left(k^{\frac{6-p}{6p}}\right)$$

for $d = 2$ *and* $d = 3$ *respectively.*

3.1.5 Numerical experiments

We have made experiments on the test system

$$\left.\begin{aligned} -\Delta u_i + \mathbf{b_i} \cdot \nabla \mathbf{u_i} + \mathbf{f_i}(\mathbf{u_1}, \ldots, \mathbf{u_l}) &= \mathbf{g_i} \\ u_{i\,|\partial\Omega} &= 0 \end{aligned}\right\} \quad (i = 1, \ldots, l) \tag{3.42}$$

on the domain $\Omega = [0,1] \times [0,1]$, where $\mathbf{b_i} = (\mathbf{1}, \mathbf{1})^{\mathrm{T}}$ for all i, and $f(\mathbf{u}) = 4\mathbf{A}\,|\mathbf{u}|^2\mathbf{u}$ where $\mathbf{A}$ is the lower triangular part of the constant 1 matrix.

The experiments were carried out in the following way:

- we used Courant elements for the FEM discretization using uniform triangle mesh with width h
- the coordinates of the exact solution were chosen among the functions of form $u(x,y) = C \cdot x(1-x)y(1-y)$ and $u(x,y) = C \cdot \sin \pi x \sin \pi y$;
- the stopping criterion was $\|F_h(\underline{u}_n) - b_h\| \leq 10^{-6}$;
- the auxiliary problems were solved with FFT;
- we used adaptive damping parameters τ_n;
- the code was written in Matlab and run on a PC.

We have run the code for the system with $l = 2, 4, 6$ equations, respectively. The results were much similar for different l with a slight increase in number of inner iterations and large increase in computing time.

We present the results in Table 1 for $l = 4$ equations, here $r_n := \|F_h(\mathbf{u}_n) - \mathbf{g}_h\|_{H_0^1}$ is the residual error at the nth outer and n_{inn} denotes the number of inner iterations. The superlinear phase of the outer DIN iteration starts around the 5th step, which is shown in Figure 1. The mesh uniform behaviour of the convergence can be observed in both the outer and inner iterations.

The CPU times are also given. These also include the time of building the finite element matrices. Since Matlab has been used, no total time-cost analysis is carried out but the CPU times only serve for illustration.

Tab. 3.1: Results for $l = 4$ equations

	$1/h = 17$		$1/h = 33$		$1/h = 49$	
n	$\|r_n\|$	n_{inn}	$\|r_n\|$	n_{inn}	$\|r_n\|$	n_{inn}
1	7.3726	1	7.4081	1	7.4151	1
2	5.3727	1	5.3940	1	5.3982	1
3	3.4515	2	3.4790	2	3.4845	2
4	1.3288	1	1.3399	2	1.3421	2
5	$6.6101 \cdot 10^{-1}$	2	$3.5355 \cdot 10^{-1}$	2	$3.5561 \cdot 10^{-1}$	2
6	$2.3429 \cdot 10^{-1}$	2	$9.2309 \cdot 10^{-2}$	5	$9.3523 \cdot 10^{-2}$	5
7	$5.7094 \cdot 10^{-2}$	5	$1.6705 \cdot 10^{-2}$	7	$1.6983 \cdot 10^{-2}$	7
8	$3.5825 \cdot 10^{-3}$	17	$2.2688 \cdot 10^{-3}$	17	$2.3033 \cdot 10^{-3}$	17
9	$3.3643 \cdot 10^{-4}$	24	$2.8591 \cdot 10^{-4}$	24	$2.9181 \cdot 10^{-4}$	24
10	$3.5510 \cdot 10^{-5}$	23	$3.7328 \cdot 10^{-5}$	37	$3.8277 \cdot 10^{-5}$	37
11	$4.4460 \cdot 10^{-6}$	41	$4.9166 \cdot 10^{-6}$	49	$5.0674 \cdot 10^{-6}$	49
CPU time(s)	$1.1822 \cdot 10^{2}$		$8.2159 \cdot 10^{2}$		$4.1348 \cdot 10^{3}$	
	$1/h = 65$		$1/h = 81$		$1/h = 97$	
n	$\|r_n\|$	n_{inn}	$\|r_n\|$	n_{inn}	$\|r_n\|$	n_{inn}
1	7.4176	1	7.4188	1	7.4194	1
2	5.3997	1	5.4004	1	5.4008	1
3	3.4865	2	3.4874	2	3.4879	2
4	1.3429	2	1.3433	2	1.3435	2
5	$3.5636 \cdot 10^{-1}$	2	$3.5670 \cdot 10^{-1}$	2	$3.5690 \cdot 10^{-1}$	2
6	$9.3961 \cdot 10^{-2}$	5	$9.4167 \cdot 10^{-2}$	5	$9.4280 \cdot 10^{-2}$	5
7	$1.7084 \cdot 10^{-2}$	7	$1.7132 \cdot 10^{-2}$	7	$1.7158 \cdot 10^{-2}$	7
8	$2.3158 \cdot 10^{-3}$	18	$2.3217 \cdot 10^{-3}$	18	$2.3249 \cdot 10^{-3}$	18
9	$2.9276 \cdot 10^{-4}$	24	$2.9376 \cdot 10^{-4}$	24	$2.9430 \cdot 10^{-4}$	24
10	$3.9288 \cdot 10^{-5}$	37	$3.9456 \cdot 10^{-5}$	37	$3.9548 \cdot 10^{-5}$	37
11	$5.2105 \cdot 10^{-6}$	49	$5.2372 \cdot 10^{-6}$	49	$5.2519 \cdot 10^{-6}$	49
CPU time(s)	$1.2864 \cdot 10^{4}$		$3.0766 \cdot 10^{4}$		$6.2980 \cdot 10^{4}$	

3.2 Semilinear elliptic interface problems

In this work we consider a class of nonlinear interface problems. Our goal is to construct a numerical method that provides superlinear convergence of the overall iteration, moreover, this convergence is mesh independent (i.e. its rate does not deteriorate as the mesh is refined). We propose an inner-outer (damped inexact Newton plus PCG) iteration for the finite element discretization of the interface problem. Our result starts from an observation used in [29]: we can recast the considered interface problem to a weak formulation, similar to that of the mixed problems. We consider matching conditions for the solution itself on the interface, i.e., the jump is allowed for the normal derivatives. It is known that the Newton method yields superlinear convergence when the exact solution of the linearized equation is given. Instead of this, one may solve the linearized equation in an inexact way, for which we apply a preconditioned conjugate gradient method. In this way we may ensure superlinear convergence of the outer Newton iterations by controlling the inaccuracy of the inner iteration, and, moreover, the inner PCG iteration also provides mesh independent superlinear convergence.

3.2.1 The interface problem

Interface problems arise in various branches of material science, biochemistry, multiphase flow etc. Such models often describe a situation when two distinct materials are involved with different conductivities or densities, another important example is from localized reaction-diffusion problems [24, 25]. Many special numerical methods have been designed for interface problems, e.g. those involving monotone iterations, see, e.g., [24, 32, 33, 34]. When one employs a fine mesh to obtain an accurate approximation, the arising large-scale system has a large condition number too, which in fact tends to infinity as the mesh parameter approaches zero.

Formulation of the problem

We consider nonlinear interface problems of the following type:

$$\begin{cases} -\operatorname{div}\,(A(x)\nabla u) + q(x,u) &= f(x) & \text{in } \Omega\setminus\Gamma, \\ [u]_\Gamma &= 0 & \text{on } \Gamma, \\ \left[A(x)\frac{\partial u}{\partial \nu}\right]_\Gamma + s(x,u) &= \gamma(x) & \text{on } \Gamma, \\ u &= g(x) & \text{on } \partial\Omega, \end{cases} \tag{3.43}$$

where $[u]_\Gamma$ and $\left[A(x)\frac{\partial u}{\partial \nu}\right]_\Gamma$ denote the jump (i.e. the difference of the limits from the two sides of the interface Γ) of u and $A(x)\frac{\partial u}{\partial \nu}$, respectively. In the case of autocatalytic

reactions, the nonlinearities often have the form $q(x,\xi) = c_1 + c_2\xi^\alpha$ (and similarly for $s(x,\xi)$).

Assumptions (A1-A4).

(A1) Ω is a bounded open domain in $\mathbb{R}^d$ ($d = 2$ or 3), the interface $\Gamma \subset \Omega$ and the boundary $\partial\Omega$ are piecewise smooth and Lipschitz continuous 1-codimensional surfaces.

(A2) $A \in L^\infty(\Omega, \mathbb{R}^{d\times d})$, for a.e. $x \in \Omega$ $A(x)$ is symmetric and it satisfies the usual condition of uniform ellipticity

$$0 < \mu_0|\xi|^2 \le \langle A(x)\xi, \xi\rangle \le \mu_1|\xi|^2$$

for some positive numbers μ_0, μ_1.

(A3) The scalar functions $q : \Omega \times \mathbb{R} \to \mathbb{R}$ and $s : \Gamma \times \mathbb{R} \to \mathbb{R}$ are measurable and bounded w.r.t. their first variable $x \in \Omega$ (resp. $x \in \Gamma$) and continuously differentiable w.r.t. their second variable $\xi \in \mathbb{R}$. Further, $f \in L^2(\Omega)$, $\gamma \in L^2(\Gamma)$ and $g \in H^1(\Omega)$.

(A4) Let $2 \le p_1$ if $d = 2$ or $2 \le p_1 < 6$ if $d = 3$, further, let $2 \le p_2$ if $d = 2$ or $2 \le p_2 < 4$ if $d = 3$. There exist constants $\alpha_1, \alpha_2, \beta_1, \beta_2 \ge 0$ such that for any $x \in \Omega$ (or $x \in \Gamma$, resp.) and $\xi \in \mathbb{R}$

$$0 \le \partial_\xi q(x,\xi) \le \alpha_1 + \beta_1|\xi|^{p_1-2}, \qquad 0 \le \partial_\xi s(x,\xi) \le \alpha_2 + \beta_2|\xi|^{p_2-2}.$$

Weak solutions

The weak solution of the above problem can be defined as a function $u^* \in H^1(\Omega)$ such that

$$\int_\Omega \Big(A(x)\nabla u^* \cdot \nabla v + q(x,u^*)v\Big)\,dx + \int_\Gamma s(x,u^*)v\,d\sigma = \int_\Omega fv\,dx + \int_\Gamma \gamma v\,d\sigma \qquad \forall v \in H^1_0(\Omega) \tag{3.44}$$

$$\text{and } u^* = g \text{ on } \partial\Omega. \tag{3.45}$$

As proved in [29], a classical solution of (3.43) is also a weak solution. Further, one has well-posedness as stated in the following theorem (using the idea of monotone operators described in e.g. [18]):

Theorem 3.14. *[29] Let Assumptions (A1-A4) hold. Then problem (3.43) has a unique weak solution u^* in $H^1(\Omega)$.*

Remark 3.15. Theorem 3.14 also holds if we allow a slightly larger class of interfaces: if the surface Γ has finitely many common points with $\partial\Omega$, then the proof in the cited paper [29] remains valid.

For simplicity, in the following we only consider homogeneous boundary conditions, i.e. $g \equiv 0$ (that is, the solution will be in $H_0^1(\Omega)$).

3.2.2 Finite element discretization

Discretization of the interface problem

We consider the finite element discretization of the interface problem (3.44). We introduce a finite element subspace $V_h = \operatorname{span}\{w_h^j, j = 1, \ldots, m\} \subset H_0^1(\Omega)$ and we seek the element $u_h \in V_h$ that satisfies

$$\int_\Omega \Big(A(x)\nabla u_h \cdot \nabla v_h + q(x, u_h)v_h\Big)\, dx + \int_\Gamma s(x, u_h)v_h\, d\sigma = \int_\Omega f v_h\, dx + \int_\Gamma \gamma v_h\, d\sigma \qquad \forall v \in V_h(\Omega). \tag{3.46}$$

This equation can be written as an equation

$$F_h(u) = f_h \tag{3.47}$$

in V_h. By similar monotonicity reasoning as in Theorem 3.14, we get

Proposition 3.16. *Under Assumptions (A1-A4), problem* (3.46) *has a unique solution* $u_h \in V_h$.

From (3.46) we are led to the problem of finding the coefficients $\mathbf{c} = (c_j)_{j=1}^m$ such that $u_h = \sum c_j w_h^j$ satisfies

$$\int_\Omega \Big(A\nabla u_h \cdot \nabla w_h^j + q(x, u_h)w_h^j\Big) + \int_\Gamma s(x, u_h)w_h^j\, d\sigma = \int_\Omega f w_h^j\, dx + \int_\Gamma \gamma w_h^j\, d\sigma, \text{ for } j = 1, \ldots, m. \tag{3.48}$$

This gives rise to a nonlinear algebraic system of the form

$$F_h(\mathbf{c}) = \mathbf{b}. \tag{3.49}$$

3.2.3 Linearization of the discretized problem

In order to use Newton iterations, we have to formulate the linearization of the nonlinear equation (3.46). At this point we introduce the inner product

$$\langle u, v\rangle_S := \int_\Omega A\nabla u \cdot \nabla v$$

on $H_0^1(\Omega)$, for which the corresponding norm $\|\cdot\|_S$ is equivalent to the standard norm of H_0^1 by the uniform ellipticity of A. Accordingly, the subspace V_h is endowed with the inherited S-inner product.

Proposition 3.17. *The function F_h defined in* (3.47) *is Gateaux differentiable, its Gateaux derivative is symmetric and satisfies for all $u_h, v_h, p_h \in V_h$*

$$\langle F'(u_h)p_h, v_h\rangle_S = \int_\Omega \Big(A\nabla p_h \cdot \nabla v_h + \partial_\xi q(x, u_h)p_h v_h\Big) + \int_\Gamma \partial_\xi s(x, u_h)p_h v_h \, d\sigma\,. \quad (3.50)$$

Proof. It follows similarly to [18, Theorem 6.2] for one equation if the Neumann boundary is replaced by the interface Γ. □

Let us denote by N_h the nonlinear part of the operator F_h, defined via

$$\langle N_h(u_h), v_h\rangle = \int_\Omega q(x, u_h)v_h \, dx + \int_\Gamma s(x, u_h)v_h \, d\sigma \qquad \forall v \in V_h(\Omega). \quad (3.51)$$

Then we obtain

$$\langle F'(u_h)p_h, v_h\rangle_S = \langle p_h + N_h'(u_h)p_h, v_h\rangle_S. \quad (3.52)$$

Finally we introduce the corresponding stiffness and mass matrices, respectively:

$$\mathbf{S_h} = \left[\langle w_h^i, w_h^j\rangle_S\right]_{i,j=1}^m, \qquad \mathbf{D_h(u_h)} = \left[\langle N_h'(u_h)w^i{}_h, w^j{}_h\rangle_{L_2}\right]_{i,j=1}^m. \quad (3.53)$$

3.2.4 The inner-outer iteration

In this section we describe the proposed inner-outer iteration and derive mesh independent superlinear convergence for both inner and outer iterations. These results require adding some additional assumptions to (A1-A4).

Assumptions (B1) or (B1').

(B1) The derivatives of q and γ w.r.t. ξ are Lipschitz continuous, that is, there are constants C_1, C_2 such that $|\partial_\xi q(x, \xi_1) - \partial_\xi q(x, \xi_2)| \le C_1|\xi_1 - \xi_2|$ for a.e. $x \in \Omega$ and $|\partial_\xi s(x, \xi_1) - \partial_\xi s(x, \xi_2)| \le C_2|\xi_1 - \xi_2|$ for a.e. $x \in \Gamma$, $\xi_1, \xi_2 \in \mathbf{R}$.

(B1') The derivatives of q and γ w.r.t. ξ are locally Lipshcitz continuous, in the sense that $|\partial_\xi q(x,\xi_1) - \partial_\xi q(x,\xi_2)| \le \left(c_1 + c_2 \max(|\xi_1|,|\xi_2|)^{p_1-3}\right)|\xi_1 - \xi_2|$ for a.e. $x \in \Omega$ for some $3 \le p_1$ (if $d=2$) or $3 \le p_1 < 6$ (if $d=3$), and $|\partial_\xi s(x,\xi_1) - \partial_\xi s(x,\xi_2)| \le \left(c_1 + c_2 \max(|\xi_1|,|\xi_2|)^{p_2-3}\right)|\xi_1 - \xi_2|$ for a.e. $x \in \Gamma$, $\xi_1, \xi_2 \in \mathbf{R}$ for some $3 \le p_2$ (if $d=2$) or $3 \le p_2 < 6$ (if $d=3$).

Outer iteration

We sum up the required properties of the discrete problem in the following proposition.

Proposition 3.18. *Let Assumptions (A1-A4,B1) hold. Then*

(1) $F_h : H_S \to H_S$ *is Gateaux differentiable;*

(2) F_h *has the form* $F_h = I_h + N_h$, *where* I_h *is the identity operator on* V_h, N_h *is also Gateaux differentiable and* $N_h'(u)$ *is symmetric for all* $u_h \in V_h$;

(3) $F_h'(v_h)$ *is regular and* $\|F_h'(v_h)w_h\| \ge \|w_h\|$ *for all* $v_h, w_h \in V_h$;

(4) if $\partial_\xi q$ *and* $\partial_\xi s$ *are bounded, then the operators* $N_h'(u_h)$ *are uniformly majorized by a symmetric compact operator* K *defined on* H_S, *in the sense that for all* $v_h \in V_h$ $\langle N'(u_h)v_h, v_h\rangle_S \le \langle Kv_h, v_h\rangle_S$, *independently of the chosen FEM subspace* V_h;

(5) in general, under assumption (A4), the operators $N_h'(u_h)$ *are only locally uniformly majorized, that is for all* $r > 0$ *there exists a selfadjoint compact operator* $K(r)$ *such that* $\langle N_h'(u_h)v_h, v_h\rangle_S \le \langle K(r)v_h, v_h\rangle_S$, $\forall v_h \in V_h$ *and for all* $\|u_h\| \le r$, *independently of the chosen FEM subspace* V_h;

(6) N_h' *is Lipschitz continous with some Lipschitz constant* L *independent of of the chosen FEM subspace* V_h;

(7) if [B1'] holds only instead of [B1] then N_h' *is only locally Lipschitz continous, with a function* $L : (0,\infty) \to (0,\infty)$ *independent of the chosen FEM subspace* V_h.

Proof. (1) It has been proved in Proposition 3.17.

(2) It follows from (3.52).

(3) Using the nonnegativity assumption on $\partial_\xi q$ and $\partial_\xi s$, we have for all $v_h, w_h \in V_h$

$$\langle F_h'(v_h)w_h, w_h\rangle_S = \int_\Omega \left(A(x)\nabla w_h \cdot \nabla w_h + \partial_\xi q(x,v_h)w_h^2\right) dx + \int_\Gamma \partial_\xi s(x,v_h)w_h^2\, d\sigma \ge \int_\Omega A(x)\nabla w_h \cdot \nabla w_h = \|w_h\|_S^2.$$

By the Cauchy-Schwarz inequality, this leads to the needed conclusion.

(4) Now $\beta_1 = \beta_2 = 0$ in assumption (A4), hence

$$\int_\Omega \partial_\xi q(x, v_h) w_h^2 \, dx + \int_\Gamma \partial_\xi s(x, v_h) w_h^2 \, d\sigma \le c_1 \int_\Omega w_h^2 \, dx + c_2 \int_\Gamma w_h^2 \, d\sigma. \tag{3.54}$$

We define the operator

$$\langle Kv, z\rangle_S \equiv \langle K_1 v, z\rangle_S + \langle K_2 v, z\rangle_S := c_1 \int_\Omega vz \, dx + c_2 \int_\Gamma vz \, d\sigma \quad (v, z \in H_0^1(\Omega)),$$

then K_1 and K_2 are bounded linear operators from $L^2(\Omega)$ resp. $L^2(\Gamma)$ to H_S. The compactness of K_1 and K_2 follows from the Rellich-Kondrachov embedding theorems in Theorem 1.19. First, because K_1 is the composition

$$H_S \overset{\text{compact embedding}}{\hookrightarrow} L^2(\Omega) \xrightarrow{K_1} H_S$$

and second, because K_2 is the composition

$$H_S \overset{\text{trace operator}}{\longrightarrow} H^{1/2}(\Gamma) \overset{\text{compact embedding}}{\hookrightarrow} L^2(\Gamma) \xrightarrow{K_2} H_S.$$

(5) By definition of $N'(u)$ we have

$$\langle N'(u)v, v\rangle_S = \int_\Omega \partial_\xi q(x, u) v^2 + \int_\Gamma \partial_\xi s(x, u) v^2. \tag{3.55}$$

We prove that both expressions have locally a compact majorant. Using assumption [A4], the Hölder inequality for the exponents $\frac{p_1-2}{p_1} + \frac{2}{p_1} = 1$ and finally the boundedness of the embedding $H_S \hookrightarrow L^{p_1}(\Omega)$ we can make estimations

$$\begin{aligned}\int_\Omega \partial_x iq(x, u) v^2 \le \int_\Omega (\alpha_1 + \beta_1 |u|^{p1-2}) v^2 \le \alpha_1 \|h\|_2^2 + \beta_1 \|u\|_{p_1} \|v^2\|_{\frac{p_1}{2}} = \\ \alpha_1 \|h\|_2^2 + \beta_1 \|u\|_{p_1} \|h\|_{p_1}^2 \le \alpha_1 \|h\|_2^2 + \beta_1 C_{p_1} \|u\| \|v\|_{p_1}^2 \end{aligned} \tag{3.56}$$

when we proved (4).

ing a suitable $0 < s < 1$ such that we have

$$H^1(\Omega) \overset{\text{compact}}{\hookrightarrow} H^s(\Omega) \overset{\text{bounded}}{\longrightarrow} L^p(\Omega).$$

Thus first we obtain the estimate

$$\|v\|_{p_1}^2 \leq C_s \|v\|_{H^s(\Omega)}^2.$$

Then we complete the proof defing the family of operators $K(r)$ for each $r > 0$ as

$$\langle K(r)u, w\rangle = \beta_1 C_{p_1} r \langle u, w\rangle_{H^s(\Omega)}.$$

Summing up the result so far, adding the fact that the embedding $H^1(\Omega) \hookrightarrow H^s(\Omega)$ is compact we have that for $\|u\| \leq r$ the operator $K(r)$ majorizes the first term in (3.55).

The second term in (3.55) can be treated in the exact same way.

(6) It follows from (7).

(7) Its proof for the term on Ω can be found in [4], hence we only need it for the term on Γ, which we now denote by S_h. Assumption (B1') implies for any (x, ξ_1) and $(x, \xi_2) \in \Omega \times \mathbf{R}$ and $\eta, \zeta \in \mathbf{R}$,

$$\left|\Big(\partial_\xi s(x,\xi_1) - \partial_\xi s(x,\xi_2)\Big)\,\eta\cdot\zeta\right| \leq \Big(c_1 + c_2\,(\max\,|\xi_1|,|\xi_2|)^{p_2-3}\Big)|\xi_1-\xi_2|\,|\eta|\,|\zeta|\,,$$

hence for all $u, v, h, z \in H_S$

$$|\langle (S_h'(u) - S_h'(v))h, z\rangle_S| = \left|\int_\Gamma \big(\partial_\xi s(x,u) - \partial_\xi s(x,v)\big)\,h\cdot z\right| d\sigma$$

$$\leq \int_\Gamma \Big(c_1 + c_2\,(\max\,|u|,|v|)^{p_2-3}\Big)|u-v|\,|h|\,|z|\,d\sigma$$

$$\leq c_1\,\|u-v\|_{L^3}\,\|h\|_{L^3}\|z\|_{L^3} \;+\; c_2\;(\max\,\|u\|_{L^{p_2}},\|v\|_{L^{p_2}})^{p_2-3}\;\|u-v\|_{L^{p_2}}\,\|h\|_{L^{p_2}}\|z\|_{L^{p_2}}$$

where $\|u\|_{L^p} \equiv \|u\|_{L^p(\Gamma)}$, and in the last estimate Hölder's inequality has been used for the cases $\frac{1}{3}+\frac{1}{3}+\frac{1}{3} = 1$ and $\frac{p_2-3}{p_2}+\frac{1}{p_2}+\frac{1}{p_2}+\frac{1}{p_2} = 1$. Then the Sobolev embedding estimate

$$\|v\|_{L^p(\Gamma)} \leq C_p \|v\|_S \quad (v \in H^1(\Omega)) \tag{3.57}$$

(valid for $3 \leq p$ if $d = 2$ and for $3 \leq p < 6$ if $d = 3$) yields

$$|\langle (S_h'(u) - S_h'(v))h, z\rangle_S|$$

$$\leq c_1\,C_3^3\,\|u-v\|_S\,\|h\|_S\|z\|_S \;+\; c_2\,C_{p_2}^{p_2}\;(\max\,\|u\|_S,\|v\|_S)^{p_2-3}\;\|u-v\|_S\,\|h\|_S\|z\|_S\,,$$

hence

$$\|S_h'(u) - S_h'(v)\| = \sup_{\substack{h,z\in H_S \\ \|h\|_S=\|z\|_S=1}} |\langle (S_h'(u) - S_h'(v))h, z\rangle_S|$$

$$\leq \left(c_1 C_3^3 + c_2 C_{p_2}^{p_2} (\max \|u\|_S, \|v\|_S)^{p_2-3}\right) \|u - v\|_S .$$

i.e. the function $L(r) := c_1 C_3^3 + c_2 C_{p_2}^{p_2} r^{p_2-3}$ can be used.

□

From item 4. of Proposition 3.18, it follows as in [4] that the jth eigenvalues of the operators $N_h'(u_h)$ are uniformly bounded by that of K, i.e. $\lambda_j(N_h'(u_h)) \leq \lambda_j(K)$. Then it obviously follows that

Proposition 3.19. *We have*

$$\frac{2}{k}\sum_{j=1}^{k} \lambda_j(N_h'(u_h)) \leq \frac{2}{k}\sum_{j=1}^{k} \lambda_j(K).$$

A similar statement holds with $K(r)$ for $\|u_h\| \leq r$ if we can only locally uniformly majorize the operators $N_h'(u_h)$; then the definition of $\lambda_j(K)$ involves an exact analogue of the expression in (32) in [4].

Now we may introduce our damped inexact Newton (DIN) method and formulate the related convergence theorem.

Theorem 3.20. *Let F_h, f_h be as defined above, then the following DIN method defined in Theorem* 1.25 *converges as stated in the theorem. And furthermore the convergence estimate is independent of the choice of V_h.*

Proof. The convergence estimates follow from [18] under the properties in Proposition 3.18. In particular, the sequence (u_{hn}) satisfies an a priori estimate $\|u_{hn}\|_S \leq R$ with some $R > 0$ independent of V_h (see, e.g. [4, Remark 4.1]), hence we have $L := L(R_0)$ as a global Lipschitz constant throughout the iteration. □

Inner iteration

For solving the inexact equalities arising in the DIN method, we use the PCG method describen on the previous chapter. Combining (3.52) and (1.4) we need to give an approximate solution to the equation

$$F_h'(u_{nh})p_h = -(F_h(u_{nh}) - f_h), \text{ which can be written as}$$

$$p_h + N_h'(u_{nh})p_h = -(F_h(u_{nh}) - f_h) = r_{nh}, \text{thus}$$

$$(\mathbf{S_h} + \mathbf{D_h}(\mathbf{u_{nh}}))\,\mathbf{p} = -\mathbf{r_n}, \tag{3.58}$$

where $p_h = \sum_j p_j w_h^j$, $r_{nh} = \sum_j r_j w_h^j$ and $\mathbf{p} = (p_j)_{j=1}^m$, $\mathbf{r_n} = (r_j)_{j=1}^m$.

This last equation is the one that we have to solve, using the preconditioner $\mathbf{S_h}$. As stated beforeit has the following convergence property for the error vector $\mathbf{e}_k = \mathbf{p}_k - \mathbf{p}$:

Theorem 3.21. *[6] The CG applied to the equation* (3.58) *yields the following convergence estimate:*

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{A_h}}}{\|\mathbf{e}_o\|_{\mathbf{A_h}}}\right)^{1/k} \leq \frac{2\|F_h'(u_{nh})^{-1}\|}{k} \sum_{j=1}^{k} \lambda_j(N_h'(u_{nh}))$$

with $\mathbf{A_h} = \mathbf{S_h} + \mathbf{D_h}(\mathbf{u_{nh}})$.

Proposition 3.19 then yields

Corollary 3.22. *We have*

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{A_h}}}{\|\mathbf{e}_o\|_{\mathbf{A_h}}}\right)^{1/k} \leq \frac{2}{k} \sum_{j=1}^{k} \lambda_j(K) =: \varepsilon_k \to 0,$$

where ε is independent of the subspace V_h used in Galerkin discretization.

Proof. By item (3) of Proposition 3.18 we have $\|F_h'(u_h)^{-1}\| \leq 1$, and the compactness of K, along with Lemma 3.9, implies that $\varepsilon_k \to 0$. □

Remark 3.23. A similar statement holds if we can only locally uniformly majorize the operators $N_h'(u_h)$. Then the sequence ε_k involves the expression mentioned after Proposition 3.19, and the above K can be replaced by $K(R)$, where $\|u_{nh}\|_S \leq R$ and (as pointed out in the proof of Theorem 3.20) this R is independent of V_h.

Explicit asymptotics

So far we have proved superlinear convergence, but as we did before we may give a more precise characterization. Namely, we can give the order of the convergence, with carefully estimating the eigenvalues $\lambda_j(N_h'(u_{nh}))$.

The minmax characterization (Proposition 1.10) gives

$$\lambda_j(N_h'(u_{nh})) = \min_{H_{n-1} \subset V_h} \max_{x \perp H_{n-1}} \frac{\langle N_h'(u_{nh}v, v\rangle}{\|v\|^2}.$$

Using the notations and the inequalities used in the proof of Proposition 3.18 we may give the upper estimate

$$\lambda_j(N_h'(u_{nh})) \leq \min_{H_{n-1} \subset V_h} \max_{v \perp H_{n-1}} \frac{c_1\|v\|^2_{L^2(\Omega)} + c_1\|v\|^2_{L^2(\Gamma)}}{\|v\|^2}.$$

in the case $\partial_\xi q(x,\xi)$ and $\partial_\xi s(x,\xi)$ are uniformly bounded. Here the first term of the min-max expression equals the jth eigenvalue of the inverse of the Laplacian.

When $\partial_\xi q$ and $\partial_\xi s$ only satisfy assumption [A4] then we have the estimate for appropriate constants C_1, C_2:

$$\lambda_j(N_h'(u_{nh})) \leq \min_{H_{n-1}\subset V_h} \max_{v\perp H_{n-1}} \frac{C_1\|u\|\|v\|^2_{L^{p_1}(\Omega)} + C_2\|u\|\|v\|^2_{L^{p_2}(\Gamma)}}{\|v\|^2}.$$

Using again the some results on Gelfand numbers [44] we may obtain explicit bounds on the eigenvaules.

Theorem 3.24. *Let Γ be a smooth, $0 < p \leq \infty$, and let*

$$1 - \frac{1}{2} - (d-1)\left(\frac{1}{2} - \frac{1}{p}\right)_+ > 0.$$

Then the Gelfand numbers of the trace operator $H^1(\Omega) \to L^p(\Omega)$ satisfy

$$c_n \sim O\left(n^{-\frac{1}{2(d-1)}}\right)$$

Hence for bounded $\partial_\xi q$,$\partial_\xi s$ we have

$$\lambda_j(N_h'(u_{nh})) \leq O(j^{-\frac{2}{d}} + j^{-\frac{1}{2(d-1)}}) = O(j^{-\frac{2}{d}})$$

and for bounded $\partial_\xi q$,$\partial_\xi s$ satisfying [A4] we have

$$\lambda_j(N_h'(u_{nh})) \leq O(j^{-\frac{1}{d}+\frac{p-2}{2p}} + j^{-\frac{1}{2(d-1)}}) = O(j^{-\frac{1}{d}+\frac{p-2}{2p}}).$$

Thus we arrive to the following explicit convergence estimates

Proposition 3.25.

- *If $\partial_\xi q(x,\xi)$ and $\partial_\xi s(x,\xi)$ are uniformly bounded then we have*

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{A_h}}}{\|\mathbf{e}_o\|_{\mathbf{A_h}}}\right)^{1/k} \leq O\left(\frac{\log k}{k}\right) \quad \text{and} \quad O\left(k^{-\frac{2}{d}+k^{-\frac{1}{d-1}}}\right)$$

  for $d = 2$ and $d = 3$ respectively.

- *If $\partial_\xi q(x,\xi)$ and $\partial_\xi s(x,\xi)$ satisfiy [A4] then we have*

$$\left(\frac{\|\mathbf{e}_k\|_{\mathbf{A_h}}}{\|\mathbf{e}_o\|_{\mathbf{A_h}}}\right)^{1/k} \leq O\left(\leq k^{-\frac{1}{d}+\frac{p-2}{2p}} + k^{-\frac{1}{d-1}}\right).$$

3.2.5 Numerical experiments

We have made experiments on some test-problems below:

- the domain was $\Omega = [0,1] \times [0,1]$, with $\Gamma = [0,1] \times \{1/2\}$
- we used Courant elements for the FEM discretization using uniform mesh with width $h = 1/N$ where N is the number of subintervals on the interval $[0,1] \times \{0\}$,
- the coordinates of the exact solutions were chosen among the functions of form

$$u(x,y) = \begin{cases} C_1 \cdot x(1-x)y(1/2-y) & \text{on } [0,1] \times (0,1/2) \\ C_2 \cdot x(1-x)(1-y)(y-1/2) & \text{on } [0,1] \times (1/2,1) \end{cases}$$

- we have chosen polynomials $q(x,\xi) := 1 + u^3$ and $s(x,\xi) := 10 + u^5$,
- the stopping criterion was $\|F_h(u_{nh}) - f_h\|_S \leq 10^{-10}$,
- the code was written in Matlab.

The results of the numerical experiments strengthen the theoretical mesh independence results and Table 3.2.

Tab. 3.2: Results for the test-problem described above

	$N = 64$		$N = 128$		$N = 192$	
n	$\|r_n\|$	n_{inn}	$\|r_n\|$	n_{inn}	$\|r_n\|$	n_{inn}
1	2.7768	1	2.7784	1	2.7787	1
2	2.5545	1	2.5562	1	2.5565	1
3	2.3322	1	2.3339	1	2.3342	1
4	2.1099	1	2.1116	1	2.1119	1
5	1.8875	1	1.8892	1	1.8895	1
6	1.6651	1	1.6668	1	1.6671	1
7	1.4426	1	1.4443	1	1.4446	1
8	1.2201	1	1.2217	1	1.2221	1
9	0.99753	1	0.99918	1	0.99949	1
10	0.77492	1	0.77657	1	0.77688	1
11	0.55228	1	0.55393	1	0.55424	1
12	0.32961	1	0.33126	1	0.33157	1
13	7.3156×10^{-3}	3	7.3741×10^{-3}	3	7.3849×10^{-3}	3
14	4.0382×10^{-6}	7	4.0782×10^{-6}	7	4.0867×10^{-6}	7
15	9.5271×10^{-12}	15	1.1658×10^{-12}	15	1.3051×10^{-12}	15

ADDENDUM

In this work we have established various superlinear convegence estimates to discretizations of elllictic equations. In final chapter we give a review and later give slight enhancements to the orders of convergence that were proven. In this work we have obtained estimates of the form

$$\frac{\|e_n\|}{\|e_0\|} \leq \varepsilon_n^n.$$

In the first chapter we used Hilbert-Schmidt methods, and showed the validity of the estimate of type

$$\varepsilon_n \leq O\left(\frac{1}{n^{1/2}}\right),$$

where the constant C_2 obtained the Hilbert-Schmidt norm of the inverse of the Laplacian for dimensions $d = 1, 2, 3$, since only for these dimensions the inverse of the Laplacian is a Hilbert-Schmidt oprerator.

In the second chapter by analysing the behavior of the Gelfand numbers of various Sobolev space embeddings we arrived to the estimate

$$\varepsilon_n \leq O\left(\frac{\log n}{n}\right) \quad \text{or} \quad O\left(n^{\alpha}\right),$$

for and appropriate α, depending on the Gelfand numbers. The derivation of this came from estimating the right hand side of the following expression

$$\varepsilon_n \leq \frac{2}{n} \sum_{j=1}^{n} \lambda_j,$$

where λ_j is a constant multiple of the jth Gelfand number. So that if the Gelfand numbers are of order j^{α} for some $\alpha < 0$ then the expression can be estimated by

$$\begin{array}{lll} O\left(\dfrac{1-n^{\alpha+1}}{n}\right) & \text{for} & \alpha < -1, \\ O\left(\dfrac{\log n}{n}\right) & \text{for} & \alpha = -1, \quad \text{and} \\ O\left(n^{\alpha}\right) & \text{for} & \alpha > -1. \end{array}$$

We remark that this second type of estimation outperform the one before. Since

in the first section the compact operator in question was the inverse of the Laplacian, by the minmax theorem its eigenvalues coincide with the Gelfand numbers of the $H_0^1(\Omega) \hookrightarrow L^2(\Omega)$ both of order $j^{-2/d}$, we arrive to the folowing improvements

$$\begin{array}{lll} O\left(\dfrac{1}{n}\right) & & \text{for } d = 1, \\ O\left(\dfrac{\log n}{n}\right) & \text{instead of} \quad O\left(\dfrac{1}{n^{1/2}}\right) & \text{for } d = 2, \\ O\left(\dfrac{1}{n^{2/3}}\right) & & \text{for } d = 3. \end{array}$$

$$\varepsilon_n^n \leq \prod_{j=1}^{n} 2\,\lambda_j.$$

Now since the values λ_j have a special form, we can easily calculate the product on the right hand side. Suppose that $\lambda_j = j^\alpha$ with $\alpha < 0$, then using the well-known lower estimate

$$n! > \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

we have that

$$\prod_{j=1}^{n} \lambda_j = n!^\alpha < C n^{\alpha/2}\left(\frac{n}{e}\right)^{n\alpha}.$$

We conclude that the following better estimates hold

$$\varepsilon_n \leq O\left(n^\alpha \cdot n^{\alpha/2n}\right).$$

For $\alpha \leq -1$ this estimation is significantly better than the one before.

SUMMARY

When solving elliptic partial differential equations numerically we arrive to ill-behaved finite dimensional algebraic equations. In case of linear equations these are linear algebraic equations with ill-conditioned matrices, this gives the direct solution methods a serious inaccuracy effect and for iteration methods a longer running time. For finite element methods the condition number of these matrices tend to infinity as the mesh gets fined, thus for iteration methods the number of iterations would tend tot infinity. So a major problem solving these equations is instability especially when using floating-point arithmetics on a computer. This problem can be weakened by various preconditioning methods. We transform the equation to a better-behaving one.

In this work we consider discretizations of some classes of elliptic equations. We show that for a well-chosen preconditioner we may achieve superlinear convergence when solving the discretized equations with convergence estimates independent of the finement of the mesh.

For nonlinear equations we show that a variant of the Newton method has superlinear convergence. For the linear equations and the linear subproblems of the Newton iterations we show superlinear convergence of the proposed preconditioned conjugate gradient method. This superlinear convergence is justified using results on the convergence properties of the conjugate gradient method for operators that are perturbations of the identity. In Chapter 1 we use the Hilbert-Schmidt norm as a tool, in chapter 2 the eigenvalues of the perturbating operators are examined to establish the converge result. In all cases we give explicit order of convergence estimates that are independent of the mesh.

In Chapter 1 we consider a class of symmetric elliptic systems and a class of symmetric nonlinear systems.

In Chapter 2 we consider a class of nonlinear nonsymmetric systems and a class of nonlinear interface problems.

For all the examinded problems numerical testresults support our results.

MAGYAR NYELVŰ ÖSSZEFOGLALÁS

Az elliptikus parciális differencálegyenletek numerikus megoldásakor előkerülő nagyméretű véges dimenziós egyenletek többnyire numerikusan rosszul viselkednek. Lineáris esetben a diszkretizációval kapott egyenletek mátrixa rosszul kondícinált, ami a

jár. Végeselem diszkretizáció esetén a kondíciószám minden határon túl nő ha a rács finomságát minden határon túl finomítjuk, ez iterációs eljárások eseten többnyire azt jelenti, hogy az adott pontosság eléréséig szükséges lépések száma is végtelenhez tart. Ez a probléma áthidalható, legalábbis enyhíthető alkalmas prekonícionálási módszert alkalmazva amellyel lényegében egy jobban kezelhető rendszert kapunk.

Jelen dolgozatban elliptikus egyenletek osztályainak diszkretizációjával foglalkozunk. Megmutatjuk, hogy egy jól megválasztott prekondicionálóval szuperlineáris konvergencia érhető el, amely konvergencia-becslése független a rács finomságától.

Nemlineáris egyenletek esetén megmutatjuk, hogy a Newton módszer megfelelő variánsa szuperlineárisan konvergál a megoldáshoz. Lineáris egyenletek illetve a Newton módszer lineáris részfeladatának esetén megmutatjuk, hogy a javasolt prekondicionált konjugált gradiens módszer is szuperlineáris konvergenciával bír. Ezen szuperlineáris konvergencia igazolásához az identitás operátor kompakt perturbációjaira vonatkozó különböző konvergenciabecsléseket használjuk. Az 1. fejezetben az alkalmazott eszköz a Hilbert-Schmidt norma, míg a 2. fejezetben a perturbáló operátor sajátértékeinek vizsgálatával kapjuk a konvergencia eredményt. Minden említett módszer esetén a konvergencia-becslések rácsfüggetlenek.

Az Az 1. fejezetben szimmetrikus lineáris es nemlineáris egyenletrendszerek egy-egy osztályát vizsgáljuk.

A 2. fejezetben nemlineáris nemszimmetrikus egyenletrendszerek egy osztályát illetve nemlineáris interface feladatok egy osztályát vizsgáljuk.

Minden említett egyenlettípusra vonatkozó konvergencia-eredményt numerikus szimulációval teszteltük.

BIBLIOGRAPHY

[1] ADAMS, R.A., *Sobolev Spaces*, Academic Press, 1975.

[2] ANTAL, I. Mesh independent superlinear convergence of the conjugate gradient method for discretized elliptic systems, Hung. Electr. Jou. Sci. HU ISSN 1418-7108: HEJ Manuscript no.: ANM-080107-A

[3] ANTAL, I. Mesh independent superliear convergence of an inner-outer iterative method for semilinear elliptic systems , NMA 2006, LNCS 4310/2007, pp. 508-515, Eds.: T. Boyanov, S. Dimova, K. Georgiev, G. Nikolov, Springer-Verlag 2007,

[4] ANTAL I., KARÁTSON J., A mesh independent superlinear algorithm for some nonlinear nonsymmetric elliptic systems, Comput. Math. Appl. 55 (2008), 2185-2196.

[5] ANTAL I., KARÁTSON J., Mesh independent superlinear convergence of an inner-outer iterative method for semilinear elliptic interface problems, J. Comp. Appl. Math. 226 (2009), 190-196.

[6] AXELSSON, O., *Iterative Solution Methods,* Cambridge University Press, 1994.

[7] AXELSSON, O., On global convergence of iterative methods, in: *Iterative solution of nonlinear systems of equations*, pp. 1-19, Lecture Notes in Math. 953, Springer, 1982.

[8] AXELSSON, O., A survey of algebraic multilevel iteration (AMLI) methods, *BIT* 43 (2003), suppl., 863–879.

[9] AXELSSON, O., KAPORIN, I., On the sublinear and superlinear rate of convergence of conjugate gradient methods. Mathematical journey through analysis, matrix theory and scientific computation (Kent, OH, 1999), *Numer. Algorithms* 25 (2000), no. 1-4, 1–22.

[10] AXELSSON, O., KARÁTSON J., Mesh independent superlinear PCG rates via compact-equivalent operators, *SIAM J. Numer. Anal.*, 45 (2007), No. 4, pp. 1495-1516 (electronic).

[11] AXELSSON, O., KARÁTSON J., Superlinearly convergent CG methods via equivalent preconditioning for nonsymmetric elliptic operators, *Numer. Math.* 99 (2004), No. 2, 197-223.

[12] BANK, R.E., ROSE, D.J., Marching algorithms for elliptic boundary value problems. I. The constant coefficient case, *SIAM J. Numer. Anal.* 14 (1977), no. 5, 792–829.

[13] CIARLET, P. G., *The finite element method for elliptic problems*, North Holland, Amsterdam, 1978.

[14] COURANT, H., HILBERT, D., *Methods of Mathematical Physics II.*, Wiley Classics Library, J. Wiley & Sons, 1989.

[15] ERN, A., GUERMOND, J-L., *Theory and Practice of Finite Elements*, Springer, 2004.

[16] FABER, V., MANTEUFFEL, T., PARTER, S.V., Necessary and sufficient conditions for the existence of a conjugate gradient method, *SIAM J. Numer. Anal.* 21 (1984), no. 2, 352–362.

[17] FABER, V., MANTEUFFEL, T., PARTER, S.V., On the theory of equivalent operators and application to the numerical solution of uniformly elliptic partial differential equations, *Adv. in Appl. Math.,* 11 (1990), 109-163.

[18] FARAGÓ I., KARÁTSON J., *Numerical solution of nonlinear elliptic problems via preconditioning operators: theory and application.* Advances in Computation, Volume 11, *NOVA Science Publishers*, New York, 2002.

[19] GOLUB, GENE H., VAN LOAN, CHARLES F., *Matrix Computations*, Johns Hopkins 1996.

[20] HACKBUSCH, W., *Multigrid methods and applications*, Springer Series in Computational Mathematics 4, Springer, Berlin, 1985.

[21] HAYES, R. M., Iterative methods of solving linear problems on Hilbert space, Contributions to the solution of systems of linear equations and the determination of eigenvalues, National Bureau of Standards Applied Mathematics Series No. 39, U. S. Government Printing Office, Washington, D. C., 1954, pp. 71-103. MR

[22] HESTENES, M. R., STIEFEL, E., Methods of Conjugate Gradients for Solving Linear Systems, J. Res. Natl. Bur. Stand. 49, 409-436 (1952).

[23] KADLEC, J., On the regularity of the solution of the Poisson problem on a domain with boundary locally similar to the boundary of a convex open set, Czechosl. Math. J., 14(89), 1964, pp. 386,393.

[24] KANDILAROV, J. D., A monotone iterative method for numerical solution of diffusion equations with nonlinear localized chemical reactions, Numerical Methods and Appllications, Lecture Notes in Computer Sciences, vol. 4310,2007, pp615-622.

[25] KANDILAROV, J. D., VULKOV, L. G.., Analysis of immersed interface difference schemes for reaction-diffusion problems with singular own sources, *Comput. Methods Appl. Math.* 3 (2003), no. 2, 253–273.

[26] KANTOROVICH L. V., AKILOV G. P., *Functional Analysis,* Pergamon Press, 1982.

[27] KARATSON, J. Mesh independent superlinear convergence of the conjugate gradient method for some equivalent self-adjoint operators, *Appl. Math.* 50 (2005), no. 3, 277–290.

[28] KARÁTSON J., FARAGÓ I., Variable preconditioning via quasi-Newton methods for nonlinear problems in Hilbert space, *SIAM J. Numer. Anal.* **41** (2003), No. 4, 1242-1262.

[29] KARÁTSON J., KOROTOV, S., Discrete maximum principles for FEM solutions of some nonlinear elliptic interface problems, Helsinki University of Technology, Institute of Mathematics, Research Report A510.

[30] KŘIŽEK, M., NEITTAANMÄKI, P., *Mathematical and numerical modelling in electrical engineering: theory and applications*, Kluwer Academic Publishers, 1996.

[31] LANCZOS, C., Solution of systems of linear equations by minimizes iterations. Journal of Research of the National Bureau of Standards, 49:33-53, 1952.

[32] LEVEQUE, R. J., LI, ZH., The immersed interface method for elliptic equations with discontinuous coefficients and singular sources, *SIAM J. Numer. Anal.* 31 (1994), no. 4, 1019–1044.

[33] LI, ZH., A fast iterative algorithm for elliptic interface problems, *SIAM J. Numer. Anal.* 35 (1998), no. 1, 230–254.

[34] LI, ZH., ITO, K., Maximum principle preserving schemes for interface problems with discontinuous coefficients, *SIAM J. Sci. Comput.* 23 (2001), no. 1, 339–361 (electronic).

[35] MORET, I., A note on the superlinear convergence of GMRES., SIAM J. Numer. Anal, vol. 34. No. 2, pp. 513-516, 1997.

[36] NEVANLINNA, O., *Convergence of iterations for linear equations,* Birkhäuser, Basel, 1993.

[37] PEDERSEN, G., K., *Analysis now*, Springer, 1989.

[38] PINKUS, A., *N-widths in approximation theory*, Springer, 1985.

[39] PLASTOCK, R., Homeomorphisms between Banach spaces, *Trans. Amer. Math. Soc.* 200 (1974), 169–183.

[40] REKTORYS, K., *The method of discretization in time and partial differential equations*, Dortrecht-Boston, Reidel, 1982.

[41] RIESZ, F., SZ.-NAGY, B., *Functional Analysis*, Courier Dover Publications, 1990.

[42] ROSSI, T., TOIVANEN, J., A parallel fast direct solver for block tridiagonal systems with separable matrices of arbitrary dimension, *SIAM J. Sci. Comput.* 20 (1999), no. 5, 1778–1796 (electronic).

[43] SAAD, Y, *Iterative Methods for Sparse Linear Systems*, SIAM, 1996.

[44] SCHNEIDER, C., Trace operators on fractalst, entropy and approximation numbers, Georgian Math. J. 18 (2011), 549-575.

[45] SWARZTRAUBER, P. N., The methods of cyclic reduction, Fourier analysis and the FACR algorithm for the discrete solution of Poisson's equation on a rectangle, *SIAM Rev.* 19 (1977), no. 3, 490–501.

[46] THOMÉE, V., *Galerkin Finite Element Methods for Parabolic Problems*, Springer, Berlin, 1997.

[47] VYBÍRAL, J., Widths of embeddings in function spaces, Journal of Complexity, 24 (2008), 545-570.

[48] RUDIN, W., *Functional Analysis*, McGraw-Hill, 1991.

[49] WINTER, R., Some superlinear convergence results for the conjugate gradient method, *SIAM J. Numer. Anal.*, 17 (1980), 14-17.

[50] ZLATEV, Z., *Computer treatment of large air pollution models*, Kluwer Academic Publishers, Dordrecht-Boston-London, 1995.

[51] ZEIDLER, E., *Nonlinear functional analysis and its applications,* Springer, 1986